有限因变量模型中的参数估计

马建军　张玉春　赵晓丽　著

電子工業出版社
Publishing House of Electronics Industry
北京・BEIJING

内 容 简 介

有限因变量模型是具有潜在变量的一大类模型的统称。根据模型中响应变量的不同可以将其分为多项选择模型、多元秩-序模型、多重二元响应模型等。本书基于逆回归方法构造了多重二元响应模型、多元秩-序模型、广义多项选择模型中回归系数的估计，证明了估计是渐近正态的，并且利用δ-方法推导了估计的渐近分布，在这个基础上构造了模型的假设检验；基于极大似然方法构造了多重二元响应Probit模型参数的渐近有效估计和多元秩-序Logit模型回归系数的估计。本书介绍的参数估计方法有效避免了维数问题。

本书适合概率论与数理统计及相关专业的科研人员阅读。

图书在版编目（CIP）数据

有限因变量模型中的参数估计/马建军, 张玉春, 赵晓丽著. —北京：电子工业出版社，2018.8

ISBN 978-7-121-35095-5

Ⅰ. ①有…　Ⅱ. ①马…　②张…　③赵…　Ⅲ. ①统计模型－参数估计　Ⅳ. ①C81

中国版本图书馆 CIP 数据核字（2018）第 216977 号

策划编辑：刘小琳
责任编辑：刘小琳　　特约编辑：许波建
印　　刷：
装　　订：
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路 173 信箱　邮编：100036
开　　本：720×1000　1/16　印张：7.5　字数：152 千字
版　　次：2018 年 8 月第 1 版
印　　次：2018 年 8 月第 1 次印刷
定　　价：39.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）88254538，liuxl@phei.com.cn。

前言

Preface

有限因变量模型（Limited Dependent Variable Models，LDV 模型）是具有潜在变量的一大类模型的统称。根据模型中响应变量的不同类型可以将其分为多项选择模型、多元秩-序模型、多重二元响应模型等。有限因变量模型是典型的高维数据模型，且模型中具有潜在变量，维数问题给有限因变量模型中的参数估计带来了很大困难。

为了解决高维数据模型中的参数估计问题，D. L. McFadden 将蒙特卡罗模拟方法应用于有限因变量模型的参数估计，相继提出了模拟矩方法、模拟极大似然方法等模拟估计方法。V. A. Hajivassiliou 在统计学手册中对有限因变量模型的模拟估计方法进行了系统的阐述。模拟估计方法计算量大、操作复杂，难以被非专业人员应用。

李克昭提出的充分降维理论方法，如切片逆回归方法（Sliced Inverse Regression）、基本海森方向（Principal Hessian Direction），可以建立降维模型以估计中心降维空间的基向量，从而在高维数据的回归问题中进行大规模数据降维。切片逆回归方法在进行数据降维时，有计算简单、估计效率高等优点，甚至有人提出，切片逆回归方法会像最小二乘法一样被广泛应用。

本书引入了数据降维的方法来构造有限因变量模型中回归系数的估计。根据有限因变量模型的特点，我们给出了一种新的逆回归方法求有限因变量模型中回归系数的估计，这种方法有效地避免了维数问题，并且有以下几个优点：

（1）构造的估计量是具有显式表达式的估计量。

（2）证明估计量具有渐近正态性。

（3）推导出回归系数估计的渐近分布。

（4）基于回归系数估计的渐近分布构造了模型相关的假设检验。

除了应用逆回归方法之外，本书中也应用极大似然方法构造了多重二元响应 Probit 模型中参数的渐近有效估计和多元秩-序 Logit 模型中回归系数的估计。在这两个模型中所构造的估计也有效地避免了维数问题。

本书的内容安排如下，第 1 章详细介绍了有限因变量模型的定义、研究方法及发展现状，第 2 章构造了多重二元响应模型回归系数的估计，第 3 章构造了多元秩-序模型回归系数的估计，第 4 章构造了广义多项选择模型回归系数的估计，第 5 章构造了多重二元响应 Probit 模型参数的渐近有效估计，第 6 章构造了多元秩-序 Logit 模型回归系数的极大似然估计。每章均研究了估计的渐近分布及假设检验问题。

本书汇聚了作者对有限因变量模型相关理论的研究成果。本书的出版还要感谢辽宁省教育厅科学研究一般项目（项目编号：LG201623）的大力支持。

由于作者水平所限，书中难免有错误和不足的地方，欢迎读者批评指正。

马建军

2018 年 6 月

目录 Contents

第 1 章　绪　论

有限因变量模型的英文全称为 Limited Dependent Variable Models，简记为 LDV 模型。LDV 模型在经济计量学中占有重要的地位，并且在很多领域有着广泛的应用，如市场调查、心理学测试、环境研究和选举学分析等。

1.1　LDV 模型的定义

1.1.1　模型的定义

定义 1.1.1　设 $p\times 1$ 向量 $\boldsymbol{Y}^*$ 与 $q\times p$ 矩阵 $\boldsymbol{X}$ 有如下关系：

$$\boldsymbol{Y}^* = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

式中，$\boldsymbol{\varepsilon}\in\mathbf{R}^p$ 为误差向量，通常 $\boldsymbol{X}$ 与 $\boldsymbol{\varepsilon}$ 独立，$\boldsymbol{\beta}\in\mathbf{R}^q$，是未知的回归系数。在一些实际问题中，$\boldsymbol{Y}^*$ 是一个潜在变量（Latent Variable），它是不可观测的，只能观测到一个与它有关的变量 $\boldsymbol{Y}$。$\boldsymbol{Y}$ 和 $\boldsymbol{Y}^*$ 之间的关系由一个映射 $\tau(\cdot)$ 确定，即

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) \tag{1.2}$$

由线性模型式（1.1）和映射式（1.2）所决定的模型称为有限因变量模型（LDV 模型）。

不同的映射 τ 决定了不同的 LDV 模型，它包含很多子模型，在本书中主要研究以下三种 LDV 模型。

1．多项选择模型

定义 1.1.2　记 $\boldsymbol{Y}^* = (y_1^*, y_2^*, \cdots, y_p^*)^{\mathrm{T}}$，若映射 τ 满足：$Y\in\{1,2,\cdots,p\}$，并且 Y 是 $\boldsymbol{Y}^*$ 中最大分量的下标，即

$$Y = \arg\max_h\{y_1^*, y_2^*, \cdots, y_p^*\} \tag{1.3}$$

这时我们称由映射式（1.3）和线性模型式（1.1）所确定的模型为多项选择模型（Multi-Nominal Choice Model，MNC 模型）。

在经济计量学中，通常称 $\boldsymbol{Y}^* = (y_1^*, y_2^*, \cdots, y_p^*)^{\mathrm{T}}$ 中的各分量为各个选择的效用函数，而 Y 是一个指示标志，表明哪一个选择的效用最高。

2．多元秩-序模型

定义 1.1.3 如果按照大小对 $\boldsymbol{Y}^*$ 中的各分量进行排序，即

$$y_{o_1}^* \geqslant y_{o_2}^* \geqslant \cdots \geqslant y_{o_p}^*$$

上述排列是向量 $\boldsymbol{Y}^*$ 中各分量的一个序，其中 $(o_1, o_2, \cdots, o_p)$ 是 $(1, 2, \cdots, p)$ 的一个置换。若定义

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (o_1, o_2, \cdots, o_p)$$

称 $\boldsymbol{Y}$ 为 $\boldsymbol{Y}^*$ 的一个序，称由此映射定义的 LDV 模型为多元秩-序模型（Multi-Variate Rank-Ordered Model，MRO 模型）。或者，下面等价的定义：令 r_j 为向量 $\boldsymbol{Y}^*$ 中小于等于第 j 个分量的分量个数，这时称 r_j 为第 j 分量的秩。同样令

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (r_1, r_2, \cdots, r_p) \tag{1.4}$$

由线性模型式（1.1）和映射式（1.4）所确定的 LDV 模型也称作多元秩-序模型。

在一些应用中，建立多元秩-序模型可能更为合理，因为在面对多个选择的时候，消费者可能会在多个选择中做出一个次序选择。相对于多项选择模型而言，多元秩-序模型包含更多的信息。

3．多重二元响应模型

定义 1.1.4 若定义映射 τ 满足 $\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (y_1, y_2, \cdots, y_p)^{\mathrm{T}}$

$$y_j = \begin{cases} 1, & y_j^* > 0 \\ 0, & y_j^* \leqslant 0 \quad j = 1, \cdots, p \end{cases} \tag{1.5}$$

式中，$\boldsymbol{Y} = (y_1, \cdots, y_p)^{\mathrm{T}}$ 是一个由 0 和 1 构成的 p 维列向量，则称线性模型式（1.1）和映射式（1.5）所确定的 LDV 模型为多重二元响应模型（Multi-Binary Response Model）。

多重二元响应模型是一种是非选择模型，对于某一个选择项或者是选择，或者是不选择，也可以描述一些二元现象是否发生，如价格的涨跌等。但多重二元响应模型在各个选择项之间不存在多项选择模型和多元秩-序模型中的大小次序关系。

以上三种模型有以下共同特点：

（1）观测不到与 $\boldsymbol{X}$ 有直接函数关系的 $\boldsymbol{Y}^*$ 的具体值，并且无法利用映射 τ 由观测值 $\boldsymbol{Y}$ 恢复 $\boldsymbol{Y}^*$ 的值[1]。在多项选择模型和多元秩-序模型中，观测到的是 $y_j^*(j = 1, 2, \cdots, p)$ 之间的一种大小关系。在多重二元响应模型中，观测到的是

$y_j^*(j=1,2,\cdots,p)$ 与 0 比较的大小关系。

（2）若不对参数进行一定的限制，那么模型是不可识别的。

对于任意一个常数 $k(k>0)$，用 k 乘以线性方程式（1.1）的两边，有 $k\boldsymbol{Y}^*=\boldsymbol{X}^{\mathrm{T}}(k\boldsymbol{\beta})+k\boldsymbol{\varepsilon}$，对于上述三种 LDV 模型有 $\boldsymbol{Y}=\tau(\boldsymbol{Y}^*)=\tau(k\boldsymbol{Y}^*)$，$\boldsymbol{Y}$ 的值保持不变。当误差项 $\boldsymbol{\varepsilon}$ 的分布假定为多元正态分布 $N(\boldsymbol{0},\boldsymbol{\Omega})$ 时，则被称为一类 Probit 模型，Probit 模型中的参数为 $(\boldsymbol{\beta},\boldsymbol{\Omega})$。对于任意常数 $k>0$，$\boldsymbol{Y}$ 在参数 $(\boldsymbol{\beta},\boldsymbol{\Omega})$ 下的分布与在参数 $(k\boldsymbol{\beta},k^2\boldsymbol{\Omega})$ 下的分布相同。因此模型中的参数是不可识别的。为了能够使模型中的参数可以识别，一般加上一些限制条件，例如，在 Probit 模型中可以令随机误差项的协方差阵 $\boldsymbol{\Omega}$ 中的第一对角线元素为 1，并称之为可识别性限[2]。

在多重二元响应模型中，如果响应变量 $y_j\ (j=1,2,\cdots,p)$ 采用如下定义：

$$\begin{cases} y_j=1, y_j^*>c_j \\ y_j=0, y_j^*\leqslant c_j \end{cases}$$

式中，$c_j\ (j=1,2,\cdots,p)$ 不全为 0，则模型中的参数在不加限制时是可以识别的。

在上述三种模型中，如果对参数没有任何限制，尽管参数是不可识别的，但是回归系数的方向是可以识别的，也就是说，回归系数可以被估计到相差一个正的常数乘积。由于以上三种模型着重于比较各选择项的大小关系，如果得到 $k\boldsymbol{\beta}\ (k>0)$ 的估计，其中 $\boldsymbol{\beta}$ 是回归系数，仍然可以建立模型，只不过是 $\boldsymbol{Y}^*$ 的尺度被放大了。由于 $y_j^*(j=1,2,\cdots,p)$ 之间被放大了相同的尺度，所以并不影响 $\boldsymbol{Y}$ 的结果。所以，如果只对回归系数感兴趣，则可以估计回归系数的方向。

1.1.2　模型的统计推断问题

以上三种 LDV 模型中观测不到 $\boldsymbol{Y}^*$，而且也不能由 $\boldsymbol{Y}$ 恢复得到 $\boldsymbol{Y}^*$，所以最小二乘法无法应用。在具有可识别限制的 LDV 模型中，目前常用的估计方法是极大似然估计。对于 Probit 模型，记 $P_y(\boldsymbol{X};\boldsymbol{\beta},\boldsymbol{\Omega})=P(\boldsymbol{Y}=\boldsymbol{y}|\ \boldsymbol{X};\boldsymbol{\beta},\boldsymbol{\Omega})$ 是给定 $\boldsymbol{X}$ 时 $\boldsymbol{Y}$ 的条件分布，令 $\{(\boldsymbol{Y}^{(i)},\boldsymbol{X}^{(i)}),i=1,2,\cdots,n\}$ 是来自 Probit 模型的一组样本，则似然函数为

$$\prod_{i=1}^{n}P_{y^{(i)}}(\boldsymbol{X}^{(i)};\boldsymbol{\beta},\boldsymbol{\Omega}) \tag{1.6}$$

当 p 较大时 $P_{y^{(i)}}(\boldsymbol{X}^{(i)};\boldsymbol{\beta},\Omega)$ 是在区域 $\left\{\boldsymbol{Y}^*:\tau(\boldsymbol{Y}^*)=\boldsymbol{Y}\right\}$ 上关于正态密度的高维积

分。当 $p \geqslant 4$ 时，一般的数值方法无法准确计算这个高维积分，目前一般利用模拟方法计算这个高维积分。由此，产生了所谓的模拟估计方法，而模拟估计方法的计算量是非常大的。在 Probit 模型中，若 $\boldsymbol{\beta}$ 为 q 维向量，扰动项 $\boldsymbol{\varepsilon}$ 的协方差阵 $\boldsymbol{\Omega}$ 为 p 阶方阵：

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ & \cdots & \cdots & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

则待估参数向量为

$$(\beta_1, \cdots, \beta_q, \sigma_{12}, \cdots, \sigma_{1p}, \sigma_{22}, \cdots, \sigma_{2p}, \cdots, \sigma_{pp})^{\mathrm{T}}$$

这是一个 $q + p(p+1)/2 - 1$ 维向量。如果 $q = 4, p = 5$，则待估参数就是一个 18 维向量。

模拟估计方法求解时需要迭代计算，每次迭代都需要计算 n 个高维积分，所以计算量非常大。

1.2 LDV 模型中参数估计问题的研究现状

对于上述三种 LDV 模型，只有当扰动项取特殊分布时才可以简化计算。McFadden[3]证明了若 ε_j $(j = 1, 2, \cdots, p)$ 独立同分布（i.i.d.），并且是极值分布时，模型可转化为 Logit 模型。对于协方差阵具有特殊形式的 LDV 模型，Heckman[4]、Bolduc[5]、Bolduc 和 Kaci[6]及 Hausman 和 Wise[7]进行了分析。

近年来，关于 LDV 模型参数估计方法的研究主要集中在模拟方法。1981 年 Lerman 和 Manski[8]首次将 Monte Calro 积分作为一种数字技术用于 LDV 模型的参数估计，Lerman 和 Manski 所使用的模拟器称为 Crude Monte Carlo（CMC）模拟器。CMC 模拟器关于 $\Pr\{\boldsymbol{D}; \boldsymbol{\mu}, \boldsymbol{\Omega}\}$ 的模拟仅仅模拟 $\boldsymbol{Y}^*$ 落入区域 $\boldsymbol{D}$ 中的频率，其中

$$\boldsymbol{D} = \{\boldsymbol{Y}^*: \boldsymbol{Y} = \tau(\boldsymbol{Y}^*)\}, \Pr\{\boldsymbol{D}; \boldsymbol{\mu}, \boldsymbol{\Omega}\} = \Pr\{\boldsymbol{Y}^* \in \boldsymbol{D} |\ \boldsymbol{X}, \boldsymbol{Y}; \boldsymbol{\mu}, \boldsymbol{\Omega}\}$$

误差项 $\boldsymbol{\varepsilon}$ 服从正态分布 $N(\boldsymbol{\mu}, \boldsymbol{\Omega})$。

CMC 模拟器的优点是计算速度较快，对于具有处理向量能力的计算软件是比较理想的。但是，CMC 模拟器关于参数不是连续的，这会导致估计的计算和渐近分布的理论存在严重的缺陷。Hendy[9]定义了 Simulation-Variance-Reduction

方法，改进了 CMC 模拟器的计算精度。

Geweke、Hajivassiliou 和 McFadden[10]及 Keane[11]提出了 GHK 模拟器，GHK 模拟器利用了重要抽样方法[12]，使得在应用中可以减小抽样的方差，并且关于参数是连续的。关于 GHK 模拟器的性质在 Borsch-Supan 和 Hajivassiliou[13]的文献中进行了讨论。Hajivassiliou 和 McFadden 考虑了利用 Gibbs 抽样技术从截断分布中抽取样本[14]，这种技术在实际应用中可以很好地收敛到真实分布，并且模拟器关于参数是连续的、可导的。

Newey 和 McFadden 提出了广义模拟矩方法[15]（Generalized Method of Simulated Moments，GMSM）。Hajivassiliou、McFadden 和 Ruudcite 讨论了多元正态的矩形概率的模拟[16]。Hajivassiliou 和 McFadden 提出了模拟得分函数的方法（MSS）[17]，并建立了模拟器的相合性和渐近正态性。Natarajan、McCulloch 和 Kiefer[18]对于多项 Probit 模型的参数估计提出了 Monte Carlo EM 方法。

利用以上模拟方法的前提条件是必须存在极大似然估计。刘金燕、徐兴忠[19]给出了多周期 Probit 模型的极大似然估计存在性的充分必要条件。孙立敏、徐兴忠[20]给出多元秩-序 Probit 模型的极大似然估计存在性的充分必要条件。赵江涛、徐兴忠给出了多项 Probit 模型的极大似然估计的充分必要条件[21]。

Nobile、McCulloch[22,23]、Ross、McCulloch[24,25]、Polson 和 Rossi[26]考虑在多项 Probit 模型的参数估计中使用先验信息，对多项 Probit 模型进行了 Bayes 分析。在可识别限制下，他们给出了 β 和 Ω 的先验分布，其中 β 的先验分布是正态分布，而 Ω 的先验分布服从逆 Wishart 分布。综合参数的先验可以得到一个后验分布，然后在后验分布中抽取样本，从而可以利用先验信息对参数进行 Bayes 估计。

关于 LDV 模型中参数估计方法的其他文献参看参考文献[27-33]。

模拟方法有以下特点：

（1）模拟极大似然估计需要较复杂的计算程序和大量的计算，同时还存在计算的收敛性和计算误差的问题。尽管目前有功能强大的计算机，但方法运用并不方便。

（2）在可识别条件限制下，用模拟方法得到的回归系数的估计仍然是一个方向的估计，而且其渐近分布和有关的假设检验、区间估计问题也没有得到完全的解决。

1.3　逆回归方法

本书的第 2～4 章将应用逆回归方法来估计回归系数，所以这里对逆回归方法进行简要介绍。

1.3.1　逆回归的总体性质

Duan 和 Li[34]提出了分片逆回归方法（Sliced Inverse Regression，SIR）。在一般回归模型中，可以用下面的条件数学期望来估计回归函数

$$\eta(\boldsymbol{x}) = E(y|\boldsymbol{x}) \tag{1.7}$$

称之为正向回归函数。当 $\boldsymbol{x}$ 的维数很高时，由于维数问题（Curse of Dimensionality）的原因，正向回归不能得到令人满意的结果。为了避开维数引起的问题可以考虑利用逆回归函数

$$\xi(y) = E(\boldsymbol{x}|y) \tag{1.8}$$

因为 y 是一维的，所以逆回归函数可以避开维数问题。逆回归估计的理论建立在下述定理的基础上。

定理 1.3.1（Duan 和 Li[34]定理 2.1）　对于一般回归模型 $y = f(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}, \varepsilon)$，如果回归变量 $\boldsymbol{x}$ 服从非退化的椭圆对称分布，那么逆回归函数式（1.8）具有下面的形式

$$\xi(y) = \boldsymbol{\mu} + \boldsymbol{\Sigma\beta\kappa}(y)$$

式中，$\boldsymbol{\mu} = E(\boldsymbol{x}), \Sigma = \mathrm{Cov}(\boldsymbol{x}), \kappa(y)$ 具有下面的形式

$$\boldsymbol{\kappa}(y) = \frac{E[(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\beta}|y]}{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma\beta}}$$

根据上述定理可以得到

$$\boldsymbol{\beta} \propto \boldsymbol{\Sigma}^{-1}[\xi(y) - \boldsymbol{\mu}] \tag{1.9}$$

式中，比例常数为 $1/\boldsymbol{\kappa}(y)$。如果 $\boldsymbol{\kappa}(y) \neq 0$，那么可以根据式（1.9）确定 $\boldsymbol{\beta}$ 的方向。

Li[35]用逆回归方法对下面的降维模型进行数据降维，即

$$y = f(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_1, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_2, \cdots, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta}_k, \varepsilon) \tag{1.10}$$

式中，$\boldsymbol{x}$ 是 $p\times 1$ 的随机向量，$\boldsymbol{\beta}_j\,(j=1,2,\cdots,k)$ 是未知回归系数，ε 与 $\boldsymbol{x}$ 独立。

当 $k<p$ 时，若能估计出 $\boldsymbol{S}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_k)$ 的一组基，便可以达到降维的目的，其中 $\boldsymbol{S}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_k)$ 表示由列向量 $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_k)$ 张成的线性空间，这个空间称作有效降维空间（EDR 空间）。

在 Li[35]的文献中放宽了对 $\boldsymbol{x}$ 分布的限制条件，即满足下面的线性条件。

A1 线性条件： 对任意 $\boldsymbol{b}\in\mathbf{R}^p$，条件期望 $E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}|\boldsymbol{x}^{\mathrm{T}}\beta_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\beta_k)$ 关于 $\boldsymbol{x}^{\mathrm{T}}\beta_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\beta_k$ 是线性的，即存在 $c_0,c_1,\cdots,c_k$，使得 $E(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{b}|\boldsymbol{x}^{\mathrm{T}}\beta_1,\cdots,\boldsymbol{x}^{\mathrm{T}}\beta_k)=c_0+c_1\boldsymbol{x}^{\mathrm{T}}\beta_1+\cdots+c_k\boldsymbol{x}^{\mathrm{T}}\beta_k$ 成立。

由于 β_j 是未知的，无法验证上述的线性条件，应用时只能要求对所有的 p 维向量都满足线性条件。当 $\boldsymbol{x}$ 的分布为上述的椭圆正态分布时满足这个条件[36-38]。而 Hall 和 Li[39]证明，当随机向量 $\boldsymbol{x}$ 的维数较大时，可以很高的概率发现随机向量 $\boldsymbol{x}$ 满足线性条件。

定理 1.3.2（Li[35]定理 3.1）　对于降维模型式（1.9），如果回归变量 $\boldsymbol{x}$ 满足条件 A1，则中心化的逆回归曲线为

$$E(\boldsymbol{x}|y)-E(\boldsymbol{x})$$

包含在由 $\boldsymbol{\Sigma}\boldsymbol{\beta}_j\,(j=1,\cdots,k)$ 的列向量所张成的子空间中，其中 $\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{x})$。

可以看到，定理 1.3.1 是定理 1.3.2 在 $k=1$ 时的特殊情况。根据定理 1.3.2，若 $E(\boldsymbol{x}|y)-E(\boldsymbol{x})$ 包含在列向量 $(\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_k)$ 所张成的子空间内，则 $\boldsymbol{\eta}_1,\cdots,\boldsymbol{\eta}_k$ 与协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的列向量所张成的线性空间正交。这样协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的 k 个非 0 特征根所对应的特征向量即 $(\eta_1,\cdots,\eta_k)$。$\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_j\,(j=1,2,\cdots,k)$ 便是向量 $(\beta_1,\cdots,\beta_k)$ 张成的 EDR 空间的一组基。只要我们能得到 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的估计，就可以得到 EDR 空间一组基的估计。

关于逆回归方法降维的有关内容参看文献[40-44]。

1.3.2　分片逆回归估计（SIR）

基于来自模型式（1.9）的一组独立同分布样本 $(y_i,\boldsymbol{x}_i)(i=1,2,\cdots,n)$，Li[35]提出了分片逆回归的估计方法（SIR）来估计 EDR 空间的一组基，步骤如下。

（1）将 $\boldsymbol{x}$ 标准化得到

$$\tilde{\boldsymbol{x}}_i=\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})\,(i=1,2,\cdots,n)$$

式中，$\bar{\boldsymbol{x}}=\dfrac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{x}_i$，$\hat{\boldsymbol{\Sigma}}=\dfrac{1}{n}\sum\limits_{i=1}^{p}(\boldsymbol{x}_i-\bar{\boldsymbol{x}})(\boldsymbol{x}_i-\bar{\boldsymbol{x}})^{\mathrm{T}}$。

（2）将 y 的值域分成 H 个片，即 $I_1,I_2,\cdots,I_H$。令 y_i 落入第 h 个片的比例为 $\hat{p}_h$，有

$$\hat{p}_h=\frac{1}{n}\sum_{i=1}^{n}\delta_h(y_i)$$

式中，$\delta_h(y_i)$ 为示性函数，当 y_i 落入第 h 片时取值为 1，否则为 0。

（3）在每个片内计算 $\tilde{\boldsymbol{x}}_i$ 的样本均值，表示为 $\hat{\boldsymbol{m}}_h(h=1,2,\cdots,H)$，有

$$\hat{\boldsymbol{m}}_h=(\frac{1}{n\hat{p}_h})\sum_{y_i\in I_h}\tilde{\boldsymbol{x}}_i$$

（4）加权协方差阵为

$$\hat{\boldsymbol{V}}=\sum_{h=1}^{H}\hat{p}_h\hat{\boldsymbol{m}}_h\hat{\boldsymbol{m}}_h'$$

式中，$\hat{\boldsymbol{V}}$ 是 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的一个样本估计。

（5）对 $\hat{\boldsymbol{V}}$ 进行特征分解，那么 k 个最大特征值所对应的特征向量便是 $\hat{\boldsymbol{\eta}}_j$ $(j=1,2,\cdots,k)$，则估计为 $\hat{\boldsymbol{\beta}}_j=\hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\eta}}_j(j=1,2,\cdots,k)$。

在上述估计方法中，由于应用分片技术对逆回归曲线进行估计，所以称之为分片逆回归方法。逆回归估计关键在于对协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的估计，在 Li 提出 SIR 方法之后，有许多统计学家研究了多种估计方法改进对协方差阵 $\mathrm{Cov}[E(\boldsymbol{x}|y)-E(\boldsymbol{x})]$ 的估计。Hsing 和 Carrol[45]考虑了 two-case 的情况，对应的协方差阵的估计为 $\hat{\boldsymbol{V}}=\dfrac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_{(2i)}-\boldsymbol{x}_{(2i-1)})(\boldsymbol{x}_{(2i)}-\boldsymbol{x}_{(2i-1)})^{\mathrm{T}}$。Zhu 和 Ng[46]研究了 H 个不同的片中所含的观测个数 c 是相等的情况，协方差阵的估计为

$$\hat{\boldsymbol{V}}=\frac{1}{n(c-1)}\sum_{h=1}^{H}\left\{\sum\sum\nolimits_{1\leqslant j<l\leqslant c}(\boldsymbol{x}_{(h,l)}-\boldsymbol{x}_{(h,j)})(\boldsymbol{x}_{(h,l)}-\boldsymbol{x}_{(h,j)})^{\mathrm{T}}\right\}$$

实际上，以上几种方法就是将 y 的值域分成若干片，然后在每个片中用起到局部平滑作用的平均值作为 $E(\boldsymbol{x}|y)$ 的估计。一些人考虑利用其他的平滑方法。Zhu 和 Fang[47]考虑了利用核估计来估计逆回归曲线。Fung、He、Liu、Shi[48]考虑了利用 B-spline basis 方法估计逆回归曲线。当 SIR 方法提出以后，人们尝试将 SIR 方法应用于不同的模型，其中 Cook [49]，以及 Cook 和 LEE[50]将 SIR 方法应用于具有二项响应变量的回归模型，但 Cook 文中的响应变量是一维的。关于 SIR 方法应用的其他文献还有[51-63]。

1.3.3 SIR 的大样本性质

SIR 估计被证明是 $\sqrt{n}$ 相合的，但是不同的估计方法的渐近性质是不相同的，其中的一些估计方法被证明是渐近正态的，而且得到了渐近方差阵的表达式。应当注意的是 $\boldsymbol{x}$ 的协方差阵 $\boldsymbol{\Sigma}=\mathrm{Cov}(\boldsymbol{x})$ 已知时与未知时的渐近分布是不同的，这一点在 Fung、He、Liu、Shi[48]的文献中有详细的叙述。

Duan 和 Li[34]证明，在一般回归模型 $y=f(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta},\varepsilon)$ 中，对于任意分片的 SIR，如果回归子 $\boldsymbol{x}$ 的分布是椭圆对称分布，则 $\boldsymbol{\beta}$ 的方向估计 $\hat{\boldsymbol{\beta}}$ 在下面的限制下是渐近正态的：

当 $\boldsymbol{\Sigma}$ 已知时， $\hat{\beta}\Sigma\hat{\beta}=1, \hat{\beta}\Sigma\hat{\beta}>0$；

当 $\boldsymbol{\Sigma}$ 未知时， $\hat{\beta}\Sigma\hat{\beta}=1, \hat{\beta}\Sigma\hat{\beta}>0$ 。

Hsing 和 Carroll[45]，Zhu 和 Ng[46]，Zhu 和 Fang[47]，以及 Fung、He、Liu、Shi[48]，Saracco[64]分别证明了他们所给出的 SIR 估计是渐近正态的。

Duan 和 Li[34]研究了渐近相对效，将 SIR 方法应用于线性回归模型，然后与最小二乘估计进行比较，发现 SIR 方法在均匀分片的情况下有很好的相对效率。

1.4　准备知识：δ-方法

为了方便读者阅读，本节简单介绍本书中所涉及的极限理论及求极限分布的 δ -方法。

1.4.1　多元 $\boldsymbol{\delta}$-方法

在很多情况下，直接求出一个函数形式统计量的精确方差或分布很困难，因此可以应用近似方法。在求这种函数估计的近似方差或分布时，一种常用的方法就是 δ -方法。

δ -方法的思想就是使用泰勒展开。记 T_n 为某个未知参数向量 $\boldsymbol{\theta}$ 相合估计的统计量，是一个已知函数，由连续映射定理可知，如果序列 T_n 概率收敛于 $\boldsymbol{\theta}$ ，且 f 在 $\boldsymbol{\theta}$ 是连续的，则 $f(T_n)$ 也依概率收敛于 $f(\boldsymbol{\theta})$ 。感兴趣的问题是如何利用统计量 T_n 的性质来获得 $f(T_n)$ 的性质。我们可以用函数 $f(T_n)$ 在 $\boldsymbol{\theta}$ 处的泰勒展开式 $f(\boldsymbol{\theta})+f'(\boldsymbol{\theta})(T_n-\boldsymbol{\theta})+\cdots$ 来近似随机向量 $f(T_n)$ ，从而获得统计量 $f(T_n)$ 的渐近性质，特别是渐近协方差和渐近分布，它可以由 $T_n-\boldsymbol{\theta}$ 的渐近分布来推导 $f(T_n)-f(\boldsymbol{\theta})$ 的渐近分布。

下面利用向量表示来讨论 δ -方法的结论。设 $T_1,\cdots,T_k$ 是随机变量，其均值分别为 $\theta_1,\cdots,\theta_k$ 。定义估计向量 $\boldsymbol{T}=(T_1,\cdots,T_k)^{\mathrm{T}}$ 和参数微量 $\boldsymbol{\theta}=(\theta_1,\cdots,\theta_k)^{\mathrm{T}}$ 。假设 $f(\boldsymbol{T})$ 是可微的，定义

$$f_i'(\boldsymbol{\theta})=\frac{\partial}{\partial t_i}g(t)\bigg|_{t_1=\theta_1,t_2=\theta_2,\cdots,t_k=\theta_k}$$

则 f 关于 $\boldsymbol{\theta}$ 的一阶泰勒展开有

$$f(\boldsymbol{t})=f(\boldsymbol{\theta})+\sum_{i=1}^{k}f_i'(\boldsymbol{\theta})(t_i-\theta_i)+\text{余项}$$

在统计应用中，可以忽略余项，因为余项是 $(t_i-\theta_i)$ 的高阶无穷小。即

$$f(\boldsymbol{t})\approx f(\boldsymbol{\theta})+\sum_{i=1}^{k}f_i'(\boldsymbol{\theta})(t_i-\theta_i)$$

对于随机向量 $\boldsymbol{T}$，两边取期望得

$$E_{\boldsymbol{\theta}}f(\boldsymbol{T})\approx f(\boldsymbol{\theta})+\sum_{i=1}^{k}f_i'(\boldsymbol{\theta})E_{\boldsymbol{\theta}}(T_i-\theta_i)=f(\boldsymbol{\theta})$$

现在可以得到 $f(\boldsymbol{T})$ 的近似方差

$$\begin{aligned}\mathrm{Var}_{\theta}[f(T)]&\approx E_{\theta}[f(\boldsymbol{T})-f(\boldsymbol{\theta})]^2\\&\approx E_{\theta}\left[\left(\sum_{i=1}^{k}f_i'(\boldsymbol{\theta})(T_i-\theta_i)\right)\right]\\&\approx\sum_{i=1}^{k}\left[f_i'(\boldsymbol{\theta})\right]^2\mathrm{Var}_{\theta}(T_i)+2\sum_{i<j}f_i'(\boldsymbol{\theta})f_j'(\boldsymbol{\theta})\mathrm{Cov}_{\theta}(T_i,T_j)\end{aligned}$$

这就是用 δ-方法获得复杂估计量的近似方差公式。由此得到下面的结果。

定理 1.4.1（多元 δ-方法） 设 $T_1(\boldsymbol{X}),\cdots,T_p(\boldsymbol{X})$ 是一个随机样本的 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)$ 的函数，满足 $E(T_i)=\theta_i$，$\mathrm{Cov}(T_i,T_j)=\sigma_{ij}$，$i,j=1,2,\cdots,p$，有

$$\boldsymbol{\Sigma}_f=\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\frac{\partial f(\boldsymbol{\theta})}{\partial\theta_i}\frac{\partial f(\boldsymbol{\theta})}{\partial\theta_j}=(\nabla_{\theta}f(\boldsymbol{\theta}))^{\mathrm{T}}\boldsymbol{\Sigma}(\nabla_{\theta}f(\boldsymbol{\theta}))$$

式中，$\nabla_{\theta}f(\boldsymbol{\theta})=\dfrac{\partial f(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}$ 是 $f(\cdot)$ 在 $\boldsymbol{\theta}$ 的梯度，且在 $\boldsymbol{\theta}$ 处取值不为 $\mathbf{0}$，则

$$\sqrt{n}\{f\left[T_1(\boldsymbol{X}),\cdots,T_p(\boldsymbol{X})\right]-f(\theta_1,\cdots,\theta_p)\}\to_d N(\mathbf{0},\boldsymbol{\Sigma}_f)$$

式中，$\to_d$ 表示依分布收敛。

推论 1.4.1 假设定理 1.4.1 中的条件成立，则 $f[T_1(\boldsymbol{X}),\cdots,T_p(\boldsymbol{X})]$ 的渐近均值和渐近方差分别是

$$E\{f[T_1(\boldsymbol{X}),\cdots,T_p(\boldsymbol{X})]\}\approx f(\theta_1,\cdots,\theta_p)$$

$$\mathrm{Var}\{f[T_1(\boldsymbol{X}),\cdots,T_p(\boldsymbol{X})]\}\approx\frac{[\nabla_{\theta}f(\boldsymbol{\theta})]^{\mathrm{T}}\boldsymbol{\Sigma}[\nabla_{\theta}f(\boldsymbol{\theta})]}{n}$$

例 1.4.1（均值比估计的矩） 设 X 和 Y 两个随机变量，分别具有非零的均值 μ_X 及 μ_Y。要估计的参数函数是 $f(\mu_X,\mu_Y)=\dfrac{\mu_X}{\mu_Y}$。容易计算得

$$\frac{\partial}{\partial \mu_X} f(\mu_X, \mu_Y) = \frac{1}{\mu_Y}，\quad \frac{\partial}{\partial \mu_Y} f(\mu_X, \mu_Y) = -\frac{\mu_X}{\mu_Y^2}$$

则有

$$E\left(\frac{X}{Y}\right) \approx \frac{\mu_X}{\mu_Y}$$

且

$$\begin{aligned}\mathrm{Var}\left(\frac{X}{Y}\right) &\approx \frac{1}{\mu_Y^2}\mathrm{Var}(X) + \frac{\mu_X^2}{\mu_Y^4}\mathrm{Var}(Y) - 2\frac{\mu_X}{\mu_Y^3}\mathrm{Cov}(X, Y) \\ &= \left(\frac{\mu_X}{\mu_Y}\right)^2 \left[\frac{\mathrm{Var}(X)}{\mu_X^2} + \frac{\mathrm{Var}(Y)}{\mu_Y^2} - \frac{2\mathrm{Cov}(X, Y)}{\mu_X \mu_Y}\right]\end{aligned}$$

从本例的结果来看，要求均值比估计的近似均值和近似方差，只要用到 X 及 Y 的均值、方差和协方差，所以说这个公式是很有用的公式。

1.4.2　向量估计函数的 δ-方法

以上讨论了多元函数的 δ-方法，实际中还有向量估计函数的情况，即估计函数是函数向量。更一般地，考虑多个估计函数，而每个函数也是多元函数的情况。设 $\boldsymbol{T_n} = (T_{n,1}, \cdots, T_{n,k})^{\mathrm{T}}$ 是一向量值的统计量，如果

$$f : R^k \mapsto R^m$$

是一个给定的映射，至少在 $\boldsymbol{\theta}$ 的邻域有定义，即

$$f(x_1, \cdots, x_k) = [f_1(x_1, \cdots, x_k), \cdots, f_m(x_1, \cdots, x_k)]^{\mathrm{T}}$$

f 在 $\boldsymbol{\theta}$ 可微的充分条件是所有偏导数 $\dfrac{\partial f_j(x)}{\partial x_i}$ 存在，且都在 $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k)^{\mathrm{T}}$ 是连续的。在任何一种情况下，全导数都可以从偏导数中找到。如果 f 是可微的，则导数 $\nabla F = \nabla F(\boldsymbol{\theta})$ 是一个 $m \times k$ 矩阵，其中

$$\nabla F(\boldsymbol{\theta}) = \begin{pmatrix} \dfrac{\partial f_1}{\partial \theta_1}(\boldsymbol{\theta}) & \cdots & \dfrac{\partial f_1}{\partial \theta_k}(\boldsymbol{\theta}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial \theta_1}(\boldsymbol{\theta}) & \cdots & \dfrac{\partial f_m}{\partial \theta_k}(\boldsymbol{\theta}) \end{pmatrix}$$

如果导数 $\nabla F(\boldsymbol{\theta})$ 在 $\boldsymbol{\theta}$ 上是连续的，则称 f 在 $\boldsymbol{\theta}$ 处连续可微。

定理 1.4.2　设 f 是定义在 $\boldsymbol{R}^k$ 一个子集上的函数向量映射，且在 $\boldsymbol{\theta}$ 是可微的。又设 $\boldsymbol{T}_n$ 是一列随机向量，取值在 f 的定义域中。当 $n \to \infty$ 时有 $\sqrt{n}(\boldsymbol{T}_n - \boldsymbol{\theta}) \to_d$

$N(\mathbf{0}, \boldsymbol{\Sigma})$，则

$$\sqrt{n}[f(\boldsymbol{T}_n) - f(\boldsymbol{\theta})] \to_d N[\mathbf{0},(\nabla F)\boldsymbol{\Sigma}(\nabla F)^{\mathrm{T}}]$$

1.5　本书提要

第 1 章介绍了三种 LDV 模型中参数的估计方法，重点介绍模拟估计方法，而模拟估计方法计算量大、计算烦琐，所以研究计算量小及计算简单、快速的估计方法是很有意义的。

如果我们只对回归系数感兴趣，可以估计回归系数的方向，逆回归方法可以用来估计这个方向。在第 2 章中研究了多重二元响应模型中回归系数的估计。首先得到了多重二元响应模型的逆回归性质，这个性质由引理 2.1.2 和定理 2.1.1 给出。利用这个性质我们可以构造回归系数的逆回归估计，不同于其他逆回归估计，这个估计是与回归系数仅相差一个正的常数乘积的估计，并且估计是显式表达，不用对一个估计矩阵求特征值和特征向量。用 δ -方法证明了估计是渐近正态的，得到了渐近协方差阵的相合估计。在此基础之上，我们讨论了基于逆回归方法的统计推断，其中包括回归系数的线性假设的双边检验、回归变量的选择、线性假设的单边检验。模拟表明基于逆回归的估计与统计推断有良好的大样本表现。

在第 3 章中讨论了多元秩-序模型逆回归方法。直接将引理 2.1.2 推广，便可以得到多元秩-序模型的逆回归性质，利用这个性质可以得到与回归系数只相差一个正常数乘积的估计。模拟表明估计有很好的效果。应用 Bootstrap 方法对回归系数的线性假设进行检验，给出了回归系数线性假设的检验统计量，并证明了检验统计量分布的 Bootstrap 估计是强相合的，由此可以进一步得到所提出的 Bootstrap 检验是渐近精确的、相合的。最后给出了模拟实验结果。

在第 4 章中我们把多项选择模型推广到更一般的形式，即将线性模型推广到一般回归模型：

$$y_j^* = g(\boldsymbol{X}_j^{\mathrm{T}} \boldsymbol{\beta}, \boldsymbol{\varepsilon})$$

式中，回归函数 g 的形式未知，我们称之为广义多项选择模型。给出了广义多项选择模型回归系数方向的估计，并证明了估计是渐近正态的。最后给出了模拟实验结果。

第 5 章考虑了多重二元响应 Probit 模型的渐近有效估计。当多重二元响应

模型中的误差项服从多元正态分布时，其每个回归方程中的误差项也服从正态分布，这样便可以得到参数的边际极大似然估计。可以证明在满足一定正则条件时，边际极大似然估计是 $\sqrt{n}$ 相合的。借助 Gibbs 抽样技术，利用模拟方法可以得到多重二元响应 Probit 模型信息阵的近似。然后利用一步迭代可以得到渐近有效估计。最后给出了模拟实验结果。由于渐近有效估计只需要一步迭代，所以计算量少，计算速度比较快。

第 6 章给出了有固定影响属性的多元秩-序模型，假定误差项独立同分布，并且服从冈贝尔 I 型极值分布，我们称之为多元秩-序 Logit 模型。推导了多元秩-序 Logit 模型中回归系数的极大似然估计，证明了极大似然估计的渐近正态，并基于渐近正态性对模型假定进行了检验。然后给出了模拟实验结果，最后针对实例进行了分析。

第 2 章　多重二元响应模型回归系数的估计

本章研究由线性模型

$$Y^* = X^{\mathrm{T}}\beta + \varepsilon \quad (X \in \mathbf{R}^{q\times p}, \beta \in \mathbf{R}^q, \varepsilon \in \mathbf{R}^p) \tag{2.1}$$

和映射 $Y = \tau(Y^*)$ 所确定的多重二元响应模型回归系数的估计问题，其中映射 τ 满足

$$y_j = \begin{cases} 1, y_j^* > 0 \\ 0, y_j^* \leqslant 0 \end{cases} \quad j = 1, \cdots, p, Y = (y_1, \cdots, y_p)^{\mathrm{T}} \tag{2.2}$$

式中， $Y = (y_1, \cdots, y_p)^{\mathrm{T}}$ 是一个由 0 和 1 构成的 p 维列向量。

在二元响应模型中，如果不对参数做可识别限制，那么只有回归系数的方向是可估的，即回归系数最多可估到相差一个正的常数乘积。由模型的特点，如果得到与回归系数只相差一个正的常数乘积的估计，便可以建立模型、进行分析和预测。

我们将逆回归方法应用于多重二元响应模型，来估计回归系数的方向。逆回归的理论要求自变量向量 X_j $(j = 1, 2, \cdots, p)$ 满足线性条件 A1。

A1 线性条件　对任意 $b \in \mathbf{R}^q$，条件期望 $E(X_j^{\mathrm{T}} b | X_j^{\mathrm{T}}\beta)$ $(j = 1, 2, \cdots, p)$ 关于 $X_j^{\mathrm{T}}\beta$ 是线性的。

本章首先得到了多重二元响应模型的逆回归性质，然后根据这条性质，不需要预先对误差项的分布做任何假定，便可以构造多重二元响应模型回归系数的估计。这是一个与回归系数只相差一个正常数乘积的估计，并且估计是显式表达，不用对一个估计矩阵进行特征分解。进一步，可以得到估计的渐近分布，从而可以对回归系数进行假设检验和区间估计。

2.1　多重二元响应模型的逆回归性质

考虑由模型式（2.1）和映射式（2.2）所确定的多重二元响应模型。令 X_j

表示矩阵 $\boldsymbol{X}$ 的第 j 列元素，$j=1,2,\cdots,p$。下面可以得到一个与定理 1.2.2 类似的结论。

引理 2.1.1（文献[35]中定理 3.1）　令 $\boldsymbol{Z}_j=\boldsymbol{X}_j-E(\boldsymbol{X}_j),\ \boldsymbol{\Sigma}_j=\mathrm{Cov}(\boldsymbol{X}_j)>0,$ $j=1,2,\cdots,p$。若 $\boldsymbol{X}_j\ (j=1,2,\cdots,p)$ 满足线性条件 A1，则有

$$E(\boldsymbol{Z}_j|y_j)\in\boldsymbol{S}(\boldsymbol{\Sigma}_j\boldsymbol{\beta}),\quad j=1,2,\cdots,p \tag{2.3}$$

式中，$\boldsymbol{S}(\boldsymbol{\Sigma}_j\boldsymbol{\beta})$ 是由矩阵 $\boldsymbol{\Sigma}_j\boldsymbol{\beta}$ 的列向量所张成的线性空间。

为了构造与回归系数只相差一个正的常数乘积的估计，需要下面的引理。

引理 2.1.2　令 W_1 和 W_2 是两个相互独立的随机变量，满足

$$V=\begin{cases}1, & W_1+W_2>0\\ 0, & W_1+W_2\leqslant 0\end{cases}$$

若 $E(W_1)$ 存在，并且 $0<p^*=P(V=1)<1$，则

$$E(W_1|V=1)\geqslant E(W_1)$$
$$E(W_1|V=0)\leqslant E(W_1)$$

另外，下面三个命题是等价的。

（1）$E(W_1|V=1)=EW_1$。

（2）$P(A_1\bigcup A_2)=0$。

其中

$$A_1=\{W_1<EW_1,-EW_1<W_2\leqslant -W_1\}$$
$$A_2=\{W_1>EW_1,-W_1<W_2\leqslant -EW_1\}$$

（3）W_1 与 V 相互独立。

证明：令 $\tilde{W}_1=W_1-EW_1$，F 和 G 分别为 W_1 和 W_2 的概率分布。

$$\begin{aligned}
& p^*E(\tilde{W}_1|V=1)-(1-p^*)E(\tilde{W}_1|V=0)\\
=\ & \int_{W_1+W_2>0}(W_1-EW_1)\mathrm{d}F(W_1)\mathrm{d}G(W_2)-\int_{W_1+W_2\leqslant 0}(W_1-EW_1)\mathrm{d}F(W_1)\mathrm{d}G(W_2)\\
=\ & \int_{-\infty}^{\infty}(W_1-EW_1)[1-G(-W_1)]\mathrm{d}F(W_1)-\int_{-\infty}^{\infty}(W_1-EW_1)G(-W_1)\mathrm{d}F(W_1)\\
=\ & \int_{-\infty}^{\infty}(W_1-EW_1)[1-2G(-W_1)]\mathrm{d}F(W_1)\\
=\ & \int_{-\infty}^{EW_1}(W_1-EW_1)[1-2G(-W_1)]\mathrm{d}F(W_1)+\int_{EW_1}^{\infty}(W_1-EW_1)[1-2G(-W_1)]\mathrm{d}F(W_1)
\end{aligned}$$

注意到式 $(W_1-EW_1)[1-2G(-W_1)]$ 在 $(-\infty,EW_1)$ 上关于 G 是单减的，在 (EW_1,∞) 上关于 G 是单增的，所以有

$$\int_{-\infty}^{EW_1}(W_1-EW_1)[1-2G(-W_1)]\mathrm{d}F(W_1)+\int_{EW_1}^{\infty}(W_1-EW_1)$$
$$[1-2G(-W_1)]\mathrm{d}F(W_1)\geqslant\int_{-\infty}^{EW_1}(W_1-EW_1)\big[1-2G(-EW_1)\big]$$
$$\mathrm{d}F(W_1)+\int_{EW_1}^{\infty}(W_1-EW_1)[1-2G(-EW_1)]\mathrm{d}F(W_1)$$
$$=[1-2G(-EW_1)]\int_{-\infty}^{\infty}(W_1-EW_1)\mathrm{d}F(W_1)=0$$

再由$0<p^*<1$和$E\tilde{W}_1=p^*E(\tilde{W}_1|V=1)+(1-p^*)E(\tilde{W}_1|V=0)=0$，可得

$$E(W_1|V=1)\geqslant EW_1, E(W_1|V=0)\leqslant EW_1$$

下面证明三个命题的等价性。

（1）$\Rightarrow$（2）

当$E(W_1|V=1)=EW_1$时，由上面的证明可得

$$\int_{-\infty}^{EW_1}(W_1-EW_1)[G(-EW_1)-G(-W_1)]\mathrm{d}F(W_1)=0$$
$$\int_{EW_1}^{+\infty}(W_1-EW_1)[G(-EW_1))-G(-W_1)]\mathrm{d}F(W_1)=0$$

因此，有

$$\int_{-\infty}^{EW_1}[G(-EW_1)-G(-W_1)]\mathrm{d}F(W_1)=-P(A_1)=0$$
$$\int_{EW_1}^{+\infty}[G(-EW_1)-G(-W_1)]\mathrm{d}F(W_1)=P(A_2)=0$$

（2）$\Rightarrow$（3）

对于任意的 Borel 集$B\subset\mathbf{R}$有

$$P(W_1\in B,V=1)=P(W_1\in B,W_1+W_2>0)$$

因为

$$\{W_1+W_2>0\}=\{W_2>-EW_1\}\cup A_2-A_1$$
$$P(V=1)=P(W_2>-EW_1)$$

得

$$\begin{aligned}&P(W_1\in B,V=1)\\&=P(W_1\in B,W_2>-EW_1)\\&=P(W_1\in B)P(W_2>-EW_1)\\&=P(W_1\in B)P(W_1+W_2>0)\\&=P(W_1\in B)P(V=1)\end{aligned}$$

（3）$\Rightarrow$（1）显然成立，证毕。

这是二元响应模型中的一条重要的性质，说明二元响应模型中关于自变量的逆回归有一定的大小关系，这有利于我们在回归分析的研究中进行一些简洁、快速的分析。

下面考虑多重二元响应模型。对于给定的$1\leqslant j\leqslant p$，令$W_1=\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta},W_2=\varepsilon_j$和 $V=y_j$。如果$\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}$和$y_j$之间不相互独立，则由引理可得

$$[E(\boldsymbol{Z}_j^{\mathrm{T}}|\ y_j=1)-E(\boldsymbol{Z}_j^{\mathrm{T}}|\ y_j=0)]\boldsymbol{\beta}>0$$

再由引理 2.1.1，存在常数k_j，使得$[E(\boldsymbol{Z}_j^{\mathrm{T}}|\ y_j=1)-E(\boldsymbol{Z}_j^{\mathrm{T}}|\ y_j=0)]\boldsymbol{\beta}=k_j\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_j\boldsymbol{\beta}>0$。又因为$\boldsymbol{\Sigma}_j$为正定矩阵，得到$k_j>0(j=1,2,\cdots,p)$。

因此得到下面的定理。

定理 2.1.1　若$0<p_j=P(y_j=1)<1,\ y_j$与$\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}(j=1,2,\cdots,p)$不相互独立，并且自变量向量$\boldsymbol{X}_j\ (j=1,2,\cdots,p)$满足线性条件 A1，则有

$$\boldsymbol{\Sigma}_j^{-1}\left[E(\boldsymbol{Z}_j|\ y_j=1)-E(\boldsymbol{Z}_j|\ y_j=0)\right]=k_j\boldsymbol{\beta} \tag{2.4}$$

式中，常数$k_j>0(j=1,2,\cdots,p)$。

2.2　多重二元响应模型回归系数的逆回归估计

基于定理 2.1.1 可以构造一个回归系数方向的估计。令$(\boldsymbol{Y}^{(i)},\boldsymbol{X}^{(i)})$ $(i=1,\cdots,n)$为来自由模型式（2.1）和映射式（2.2）所确定的多重二元响应模型的一组 i.i.d.样本，即

$$\boldsymbol{\mu}_{j1}=E(\boldsymbol{Z}_j|\ y_j=1),\quad \boldsymbol{\mu}_{j0}=E(\boldsymbol{Z}_j|\ y_j=0),\quad \boldsymbol{\mu}_j=E(\boldsymbol{X}_j)$$

条件期望$\boldsymbol{\mu}_{j0}$和$\boldsymbol{\mu}_{j1}$的估计为

$$\hat{\boldsymbol{\mu}}_{j0}=\frac{\sum_{i=1}^{n}\boldsymbol{Z}_j^{(i)}(1-y_j^{(i)})}{\sum_{i=1}^{n}(1-y_j^{(i)})}$$

$$\hat{\boldsymbol{\mu}}_{j1}=\frac{\sum_{i=1}^{n}\boldsymbol{Z}_j^{(i)}y_j^{(i)}}{\sum_{i=1}^{n}y_j^{(i)}},\ j=1,2,\cdots,p$$

其中

$$\boldsymbol{Z}_j^{(i)} = (\boldsymbol{X}_j^{(i)} - \hat{\boldsymbol{\mu}}_j)$$

$$\hat{\boldsymbol{\mu}}_j = \frac{1}{n}\sum_{i=1}^{n} X_j^{(i)},\ j = 1, 2, \cdots, p$$

则回归系数 $\boldsymbol{\beta}$ 方向的一个估计由下面的式子给出。当 $\boldsymbol{\Sigma}_j$ $(j=1,\cdots,p)$ 已知时

$$\tilde{\gamma}_j = \boldsymbol{\Sigma}_j^{-1}(\hat{\boldsymbol{\mu}}_{j1} - \hat{\boldsymbol{\mu}}_{j0}),\ j = 1, 2, \cdots, p$$

当 $\boldsymbol{\Sigma}_j\,(j=1, 2,\cdots, p)$ 未知时

$$\hat{\gamma}_j = \hat{\boldsymbol{\Sigma}}_j^{-1}(\hat{\boldsymbol{\mu}}_{j1} - \hat{\boldsymbol{\mu}}_{j0}),\ j = 1, 2, \cdots, p$$

其中，$\hat{\boldsymbol{\Sigma}}_j$ 为 $\boldsymbol{\Sigma}_j$ 的矩估计

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_j^{(i)} - \boldsymbol{\mu}_j)(\boldsymbol{X}_j^{(i)} - \boldsymbol{\mu}_j)^{\mathrm{T}}$$

这样的估计一共有 p 个。为了利用全部数据中的信息来估计 $\boldsymbol{\beta}$，由于所有的 $p_j > 0$，利用这 p 个估计的算术平均作为 $\boldsymbol{\beta}$ 方向的估计：

$$\tilde{\gamma} = \frac{1}{p}\sum_{j=1}^{p}\boldsymbol{\Sigma}_j^{-1}(\hat{\boldsymbol{\mu}}_{j1} - \hat{\boldsymbol{\mu}}_{j0}),\ \boldsymbol{\Sigma}_j\,(j=1, 2,\cdots, p)\text{已知。} \tag{2.5}$$

$$\hat{\gamma} = \frac{1}{p}\sum_{j=1}^{p}\hat{\boldsymbol{\Sigma}}_j^{-1}(\hat{\boldsymbol{\mu}}_{j1} - \hat{\boldsymbol{\mu}}_{j0}),\ \boldsymbol{\Sigma}_j\,(j=1, 2,\cdots, p)\text{未知。} \tag{2.6}$$

实际上，$\hat{\gamma}$ 就是 $k\boldsymbol{\beta}$ 的估计，其中 $k = \frac{1}{p}\sum_{j=1}^{p}k_j$ 。

2.3 估计的渐近正态性

在这节中，我们证明所给出的估计是渐近正态的，并导出估计的渐近分布。$\boldsymbol{\Sigma}_j\,(j=1,2,\cdots,p)$ 已知和未知时估计的渐近分布是不同的，这个问题在 Fung、He、Liu 与 Shi[47]的论文中有比较详细的讨论。

定理 2.3.1 在引理 2.1.1 的条件下，如果 $\boldsymbol{\Sigma}_j\,(j=1,\cdots,p)$ 已知，则有

$$\sqrt{n}(\tilde{\gamma} - \gamma) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega}_{\gamma})$$

其中

$$\boldsymbol{\Omega}_{\gamma} = \frac{1}{p^2}\boldsymbol{\Gamma}\boldsymbol{R}_3\boldsymbol{R}_2\boldsymbol{R}_1\boldsymbol{\Omega}_1\boldsymbol{R}_1^{\mathrm{T}}\boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}} \tag{2.7}$$

$$\boldsymbol{\Omega}_1=\mathrm{Cov}(y_1,\cdots,y_p,\boldsymbol{X}_1^{\mathrm{T}}y_1,\cdots,\boldsymbol{X}_p^{\mathrm{T}}y_p,\boldsymbol{X}_1^{\mathrm{T}}(1-y_1),\cdots,\boldsymbol{X}_p^{\mathrm{T}}(1-y_p),\boldsymbol{X}_1^{\mathrm{T}},\cdots,\boldsymbol{X}_p^{\mathrm{T}}) \tag{2.8}$$

$$\boldsymbol{R}_1=\begin{bmatrix}\boldsymbol{I}_p & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq}\\ \boldsymbol{D}_1 & \boldsymbol{I}_{pq} & \boldsymbol{O}_{pq} & \boldsymbol{D}_2\\ -\boldsymbol{D}_1 & \boldsymbol{O}_{pq} & \boldsymbol{I}_{pq} & \boldsymbol{D}_3\end{bmatrix}$$

其中，$\boldsymbol{O}_m$ 为一个 $m\times m$ 的零矩阵，$\boldsymbol{O}_{m,n}$ 为一个 $m\times n$ 的零矩阵，$\boldsymbol{I}_m$ 为 $m\times m$ 单位阵，且有

$$\boldsymbol{D}_1=\begin{bmatrix}-\mu_1 & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & -\mu_2 & \cdots & \boldsymbol{0}\\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\mu_p\end{bmatrix}$$

$$\boldsymbol{D}_2=\begin{bmatrix}-p_1\boldsymbol{I}_q & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & -p_2\boldsymbol{I}_q & \cdots & \boldsymbol{0}\\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -p_p\boldsymbol{I}_q\end{bmatrix} \tag{2.9}$$

$$\boldsymbol{D}_3=\begin{bmatrix}-(1-p_1)\boldsymbol{I}_q & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & -(1-p_2)\boldsymbol{I}_q & \cdots & \boldsymbol{0}\\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -(1-p_p)\boldsymbol{I}_q\end{bmatrix}$$

另外，$\boldsymbol{R}_2=\begin{bmatrix}\boldsymbol{E}_1 & \boldsymbol{E}_2 & \boldsymbol{O}_{pq}\\ \boldsymbol{E}_3 & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{E}_4\end{bmatrix}$。

其中，有

$$\boldsymbol{E}_1=\begin{bmatrix}-\dfrac{\mu_{11}}{p_1} & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & -\dfrac{\mu_{21}}{p_2} & \cdots & \boldsymbol{0}\\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\dfrac{\mu_{p1}}{p_p}\end{bmatrix}$$

$$\boldsymbol{E}_2 = \begin{bmatrix} \dfrac{\boldsymbol{I}_q}{p_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \dfrac{\boldsymbol{I}_q}{p_2} & \cdots & \boldsymbol{0} \\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \dfrac{\boldsymbol{I}_q}{p_p} \end{bmatrix} \tag{2.10}$$

$$\boldsymbol{E}_3 = \begin{bmatrix} \dfrac{\boldsymbol{\mu}_{10}}{1-p_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \dfrac{\boldsymbol{\mu}_{20}}{1-p_2} & \cdots & \boldsymbol{0} \\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \dfrac{\boldsymbol{\mu}_{p0}}{1-p_p} \end{bmatrix}$$

$$\boldsymbol{E}_4 = \begin{bmatrix} \dfrac{\boldsymbol{I}_q}{1-p_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \dfrac{\boldsymbol{I}_q}{1-p_2} & \cdots & \boldsymbol{0} \\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \dfrac{\boldsymbol{I}_q}{1-p_p} \end{bmatrix}$$

$$\boldsymbol{R}_3 = [\boldsymbol{I}_{pq}, -\boldsymbol{I}_{pq}]$$

$$\boldsymbol{\Gamma} = \left[\boldsymbol{\Sigma}_1^{-1}, \cdots, \boldsymbol{\Sigma}_p^{-1}\right]$$

证明：令

$$\begin{aligned} \boldsymbol{U}_1^{(i)} = [& y_1^{(i)}, \cdots, y_p^{(i)}, \boldsymbol{X}_1^{(i)\mathrm{T}} y_1^{(i)}, \cdots, \boldsymbol{X}_p^{(i)\mathrm{T}} y_p^{(i)}, \\ & \boldsymbol{X}_1^{(i)\mathrm{T}}(1-y_1^{(i)}), \cdots, \boldsymbol{X}_p^{(i)\mathrm{T}}(1-y_p^{(i)}), \boldsymbol{X}_1^{(i)\mathrm{T}}, \cdots, \boldsymbol{X}_p^{(i)\mathrm{T}}]^{\mathrm{T}} \end{aligned} \tag{2.11}$$

$$\bar{\boldsymbol{U}}_1 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{U}_1^{(i)} \tag{2.12}$$

则 $\boldsymbol{U}_1^{(i)}$ $(i=1, 2, \cdots, n)$ i.i.d.，期望 $E(\boldsymbol{U}_1)$ 为

$$\begin{aligned} E(\boldsymbol{U}_1) = [& p_1, \cdots, p_p, p_1(\boldsymbol{\mu}_{11}^{\mathrm{T}} + \boldsymbol{\mu}_1^{\mathrm{T}}), \cdots, p_p(\boldsymbol{\mu}_{p1}^{\mathrm{T}} + \boldsymbol{\mu}_p^{\mathrm{T}}), \\ & (1-p_1)(\boldsymbol{\mu}_{10}^{\mathrm{T}} + \boldsymbol{\mu}_1^{\mathrm{T}}), \cdots, (1-p_p)(\boldsymbol{\mu}_{p0}^{\mathrm{T}} + \boldsymbol{\mu}_p^{\mathrm{T}}), \boldsymbol{\mu}_1^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_p^{\mathrm{T}}]^{\mathrm{T}} \end{aligned}$$

应用中心极限定理得

$$\sqrt{n}[\bar{\boldsymbol{U}}_1 - E(\boldsymbol{U}_1)] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_1)$$

其中

$$\boldsymbol{\Omega}_1 = \mathrm{Cov}(y_1, \cdots, y_p, \boldsymbol{X}_1{}^{\mathrm{T}} y_1, \cdots, \boldsymbol{X}_p{}^{\mathrm{T}} y_p, \boldsymbol{X}_1{}^{\mathrm{T}}(1-y_1), \cdots, \\ \boldsymbol{X}_p{}^{\mathrm{T}}(1-y_p), \boldsymbol{X}_1{}^{\mathrm{T}}, \cdots, \boldsymbol{X}_p{}^{\mathrm{T}})$$

下面考虑函数 $\boldsymbol{g}_1$：

$$\begin{array}{ccc} \mathbf{R}^{p+pq+pq+pq} & \to & \mathbf{R}^{p+pq+pq} \\ \boldsymbol{u} = (\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}}, \boldsymbol{c}_1^{\mathrm{T}}, \cdots, \boldsymbol{c}_p^{\mathrm{T}}, & \mapsto & [\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}_1^{\mathrm{T}} - a_1 \boldsymbol{d}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}} - a_p \boldsymbol{d}_p^{\mathrm{T}}, \\ \boldsymbol{d}_1^{\mathrm{T}}, \cdots, \boldsymbol{d}_p^{\mathrm{T}})^{\mathrm{T}} & & \boldsymbol{c}_1^{\mathrm{T}} - (1-a_1)\boldsymbol{d}_1^{\mathrm{T}}, \cdots, \boldsymbol{c}_p^{\mathrm{T}} - (1-a_p)\boldsymbol{d}_p^{\mathrm{T}}]^{\mathrm{T}} \end{array}$$

其中 $\boldsymbol{a} = (a_1, \cdots, a_p)^{\mathrm{T}} \in \mathbf{R}^p, \boldsymbol{b}_j, \boldsymbol{c}_j, \boldsymbol{d}_j \in \mathbf{R}^q$。

令 $\boldsymbol{U}_2 = \boldsymbol{g}_1(\bar{\boldsymbol{U}}_1)$，则

$$\begin{aligned} E(\boldsymbol{U}_2) &= \boldsymbol{g}_1[E(\boldsymbol{U}_1)] \\ &= [p_1, \cdots, p_p, p_1\boldsymbol{\mu}_{11}^{\mathrm{T}}, \cdots, p_p\boldsymbol{\mu}_{p1}^{\mathrm{T}}, (1-p_1)\boldsymbol{\mu}_{10}^{\mathrm{T}}, \cdots, (1-p_p)\boldsymbol{\mu}_{p0}^{\mathrm{T}}]^{\mathrm{T}} \end{aligned}$$

利用 δ-方法可得

$$\sqrt{n}[\boldsymbol{U}_2 - E(\boldsymbol{U}_2)] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_2)$$

其中协方差阵为

$$\boldsymbol{\Omega}_2 = \boldsymbol{R}_1 \boldsymbol{\Omega}_1 \boldsymbol{R}_1^{\mathrm{T}}$$

$\boldsymbol{R}_1 = \left.\dfrac{\partial \boldsymbol{g}_1}{\partial \boldsymbol{u}^{\mathrm{T}}}\right|_E$，是 $\dfrac{\partial \boldsymbol{g}_1}{\partial \boldsymbol{u}^{\mathrm{T}}}$ 在其各分量期望处的取值，有

$$\boldsymbol{R}_1 = \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} \\ \boldsymbol{D}_1 & \boldsymbol{I}_{pq} & \boldsymbol{O}_{pq} & \boldsymbol{D}_2 \\ -\boldsymbol{D}_1 & \boldsymbol{O}_{pq} & \boldsymbol{I}_{pq} & \boldsymbol{D}_3 \end{bmatrix}$$

其中，$\boldsymbol{O}_m$ 为一个 $m \times m$ 的零矩阵，$\boldsymbol{O}_{m,n}$ 为一个 $m \times n$ 的零矩阵，$\boldsymbol{I}_m$ 为 $m \times m$ 的单位阵，并且

$$\boldsymbol{D}_1 = \begin{bmatrix} -\mu_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & -\mu_2 & \cdots & \mathbf{0} \\ & \cdots & \cdots & \\ \mathbf{0} & \mathbf{0} & \cdots & -\mu_p \end{bmatrix}$$

$$\boldsymbol{D}_2=\begin{bmatrix} -p_1\boldsymbol{I}_q & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & -p_2\boldsymbol{I}_q & \cdots & \boldsymbol{0} \\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -p_p\boldsymbol{I}_q \end{bmatrix}$$

$$\boldsymbol{D}_3=\begin{bmatrix} -(1-p_1)\boldsymbol{I}_q & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & -(1-p_2)\boldsymbol{I}_q & \cdots & \boldsymbol{0} \\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -(1-p_p)\boldsymbol{I}_q \end{bmatrix}$$

下面考虑函数 $\boldsymbol{g}_2$：

$$\begin{array}{ccc} \mathbf{R}^{p+pq+pq} & \to & \mathbf{R}^{pq+pq} \\ \boldsymbol{u}=(\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\boldsymbol{c}_1^{\mathrm{T}},\cdots,\boldsymbol{c}_p^{\mathrm{T}})^{\mathrm{T}} & \mapsto & (\dfrac{\boldsymbol{b}_1^{\mathrm{T}}}{a_1},\cdots,\dfrac{\boldsymbol{b}_p^{\mathrm{T}}}{a_p},\dfrac{\boldsymbol{c}_1^{\mathrm{T}}}{1-a_1},\cdots,\dfrac{\boldsymbol{c}_p^{\mathrm{T}}}{1-a_p})^{\mathrm{T}} \end{array}$$

其中，$\boldsymbol{a}=(a_1,\cdots,a_p)^{\mathrm{T}}\in\mathbf{R}^p,\boldsymbol{b}_j,\boldsymbol{c}_j\in\mathbf{R}^q$。

令

$$\boldsymbol{U}_3=\boldsymbol{g}_2(\boldsymbol{U}_2)=\left(\hat{\boldsymbol{\mu}}_{11}^{\mathrm{T}},\cdots,\hat{\boldsymbol{\mu}}_{p1}^{\mathrm{T}},\hat{\boldsymbol{\mu}}_{10}^{\mathrm{T}},\cdots,\hat{\boldsymbol{\mu}}_{p0}^{\mathrm{T}}\right)^{\mathrm{T}}$$

则

$$E(\boldsymbol{U}_3)=\boldsymbol{g}_2[E(\boldsymbol{U}_2)]=\left(\boldsymbol{\mu}_{11}^{\mathrm{T}},\cdots,\boldsymbol{\mu}_{p1}^{\mathrm{T}},\boldsymbol{\mu}_{10}^{\mathrm{T}},\cdots,\boldsymbol{\mu}_{p0}^{\mathrm{T}}\right)^{\mathrm{T}}$$

我们得到

$$\sqrt{n}[\boldsymbol{U}_3-E(\boldsymbol{U}_3)]\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{\Omega}_3)$$

其中协方差阵 $\boldsymbol{\Omega}_3$ 为

$$\boldsymbol{\Omega}_3=\boldsymbol{R}_2\boldsymbol{\Omega}_2\boldsymbol{R}_2^{\mathrm{T}}$$

$$\boldsymbol{R}_2=\begin{bmatrix} \boldsymbol{E}_1 & \boldsymbol{E}_2 & \boldsymbol{O}_{pq} \\ \boldsymbol{E}_3 & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{E}_4 \end{bmatrix}$$

矩阵 $\boldsymbol{R}_2$ 中的 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$、$\boldsymbol{E}_3$、$\boldsymbol{E}_4$ 为

$$\boldsymbol{E}_1=\begin{bmatrix} -\dfrac{\boldsymbol{\mu}_{11}}{p_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & -\dfrac{\boldsymbol{\mu}_{21}}{p_2} & \cdots & \boldsymbol{0} \\ & \cdots & \cdots & \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\dfrac{\boldsymbol{\mu}_{p1}}{p_p} \end{bmatrix}$$

$$\boldsymbol{E}_2 = \begin{bmatrix} \frac{\boldsymbol{I}_q}{p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{\boldsymbol{I}_q}{p_2} & \cdots & \mathbf{0} \\ & \cdots & \cdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{\boldsymbol{I}_q}{p_p} \end{bmatrix}$$

$$\boldsymbol{E}_3 = \begin{bmatrix} \frac{\boldsymbol{\mu}_{10}}{1-p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{\boldsymbol{\mu}_{20}}{1-p_2} & \cdots & \mathbf{0} \\ & \cdots & \cdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{\boldsymbol{\mu}_{p0}}{1-p_p} \end{bmatrix}$$

$$\boldsymbol{E}_4 = \begin{bmatrix} \frac{\boldsymbol{I}_q}{1-p_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \frac{\boldsymbol{I}_q}{1-p_2} & \cdots & \mathbf{0} \\ & \cdots & \cdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \frac{\boldsymbol{I}_q}{1-p_p} \end{bmatrix}$$

下面定义函数 $\boldsymbol{g}_3$：

$$\begin{array}{ccc} \mathbf{R}^{pq+pq} & \to & \mathbf{R}^{pq} \\ \boldsymbol{u} = (\boldsymbol{b}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}}, \boldsymbol{c}_1^{\mathrm{T}}, \cdots, \boldsymbol{c}_p^{\mathrm{T}})^{\mathrm{T}} & \mapsto & (\boldsymbol{b}_1^{\mathrm{T}} - \boldsymbol{c}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}} - \boldsymbol{c}_p^{\mathrm{T}})^{\mathrm{T}} \end{array}$$

其中，$\boldsymbol{b}_j, \boldsymbol{c}_j \in \mathbf{R}^q$。

令

$$\boldsymbol{U}_4 = \boldsymbol{g}_3(\boldsymbol{U}_3) = (\hat{\boldsymbol{\mu}}_{11}^{\mathrm{T}} - \hat{\boldsymbol{\mu}}_{10}^{\mathrm{T}}, \cdots, \hat{\boldsymbol{\mu}}_{p1}^{\mathrm{T}} - \hat{\boldsymbol{\mu}}_{p0}^{\mathrm{T}})^{\mathrm{T}}$$

则

$$E(\boldsymbol{U}_4) = \boldsymbol{g}_3[E(\boldsymbol{U}_3)] = (\boldsymbol{\mu}_{11}^{\mathrm{T}} - \boldsymbol{\mu}_{10}^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_{p1}^{\mathrm{T}} - \boldsymbol{\mu}_{p0}^{\mathrm{T}})^{\mathrm{T}}$$

由 δ-方法得

$$\sqrt{n}[\boldsymbol{U}_4 - E(\boldsymbol{U}_4)] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_4)$$

其中

$$\boldsymbol{\Omega}_4 = \boldsymbol{R}_3 \boldsymbol{\Omega}_3 \boldsymbol{R}_3^{\mathrm{T}}$$

$$\boldsymbol{R}_3 = [\boldsymbol{I}_{pq}, -\boldsymbol{I}_{pq}]$$

令 $\boldsymbol{\Gamma} = \left[\boldsymbol{\Sigma}_1^{-1}, \cdots, \boldsymbol{\Sigma}_p^{-1}\right]$，则

$$\tilde{\gamma} = \frac{1}{p} \boldsymbol{\Gamma} \boldsymbol{U}_4$$

$$\gamma = \frac{1}{p} \boldsymbol{\Gamma} E(\boldsymbol{U}_4)$$

因此，可以得到结论

$$\sqrt{n}(\tilde{\gamma} - \gamma) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega}_\gamma)$$

其中

$$\boldsymbol{\Omega}_\gamma = \frac{1}{p^2} \boldsymbol{\Gamma} \boldsymbol{\Omega}_4 \boldsymbol{\Gamma}^{\mathrm{T}}$$

证毕。

上面的证明应用 δ-方法推导了逆回归估计的渐近分布，目前有许多文献讨论过逆回归的渐近分布的问题，其中包括参考文献[45-47,51]。但是，在这些方法中，都需要响应变量与自变量有一种直接的数量关系，而在多重二元响应模型中与自变量有直接数量关系的是一个潜在变量，无法观测到。所以这里使用 δ-方法，尽管步骤较为烦琐，还是比较适合的。

在实际应用中按照下面的步骤可以得到 $\boldsymbol{\Omega}_\gamma$ 的一个相合估计。

（1）首先得到 $\boldsymbol{\Omega}_1$ 的估计 $\hat{\boldsymbol{\Omega}}_1$：

$$\hat{\boldsymbol{\Omega}}_1 = \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{U}_1^{(i)} - \bar{\boldsymbol{U}}_1\right)\left(\boldsymbol{U}_1^{(i)} - \bar{\boldsymbol{U}}_1\right)^{\mathrm{T}}$$

其中，$\boldsymbol{U}_1^{(i)}$ 和 $\bar{\boldsymbol{U}}_1$ 由式（2.11）和式（2.12）确定。

（2）注意到在定理 2.3.1 的证明过程中的矩阵 $\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 只涉及 p_j、$\boldsymbol{\mu}_{j1}$、$\boldsymbol{\mu}_{j0}$ 和 $\boldsymbol{\mu}_j$，而 $\boldsymbol{R}_3$ 和 $\boldsymbol{\Gamma}$ 是已知的，我们只需要用矩估计 $\hat{p}_j$、$\hat{\boldsymbol{\mu}}_{j1}$、$\hat{\boldsymbol{\mu}}_{j0}$、$\hat{\boldsymbol{\mu}}_j$ 代替 p_j、$\boldsymbol{\mu}_{j1}$、$\boldsymbol{\mu}_{j0}$、$\boldsymbol{\mu}_j$，便可以得到 $\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 的相合估计。

（3）将 $\boldsymbol{R}_1$ 和 $\boldsymbol{R}_2$ 的相合估计与 $\boldsymbol{R}_3$、$\boldsymbol{\Gamma}$ 代入式（2.7）中，便可以得到 $\boldsymbol{\Omega}_\gamma$ 的相合估计。

下面证明当 $\boldsymbol{\Sigma}_j\ (j=1,\cdots,p)$ 未知，并且由数据得到其相合估计时，所构造的估计 $\hat{\gamma}$ 的渐近正态性。

定理 2.3.2　若 $\boldsymbol{X}_j(j=1, 2,\cdots, p)$ 存在有限的四阶矩。在引理 2.1.1 的条件下，如果 $\varSigma_j(j=1,\cdots, p)$ 未知，并且由数据估计所得，则有

$$\sqrt{n}(\hat{\gamma}-\gamma)\xrightarrow{d}N(\mathbf{0}, \boldsymbol{\varOmega}_\gamma^*)$$

其中，渐近协方差阵 $\boldsymbol{\varOmega}_\gamma^*$ 的表达式由下面的证明过程和式（2.13）给出。

其中

$$\boldsymbol{\varOmega}_\gamma^*=\frac{1}{p^2}\boldsymbol{R}_5^*\boldsymbol{R}_4^*\boldsymbol{R}_3^*\boldsymbol{R}_2^*\boldsymbol{R}_1^*\boldsymbol{\varOmega}_1^*\boldsymbol{R}_1^{*\mathrm{T}}\boldsymbol{R}_2^{*\mathrm{T}}\boldsymbol{R}_3^{*\mathrm{T}}\boldsymbol{R}_4^{*\mathrm{T}}\boldsymbol{R}_5^{*\mathrm{T}} \tag{2.13}$$

$$\begin{aligned}\boldsymbol{\varOmega}_1^*=\mathrm{Cov}[&y_1,\cdots,y_p,\boldsymbol{X}_1^{\mathrm{T}}y_1,\cdots,\boldsymbol{X}_p^{\mathrm{T}}y_p,\boldsymbol{X}_1^{\mathrm{T}}(1-y_1),\cdots,\boldsymbol{X}_p^{\mathrm{T}}(1-y_p),\\&\boldsymbol{X}_1^{\mathrm{T}},\cdots,\ \boldsymbol{X}_p^{\mathrm{T}},\mathrm{vec}(\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{T}})^{\mathrm{T}},\cdots,\mathrm{vec}(\boldsymbol{X}_p\boldsymbol{X}_p^{\mathrm{T}})^{\mathrm{T}}]^{\mathrm{T}}\end{aligned} \tag{2.14}$$

$$\boldsymbol{R}_1^*=\begin{bmatrix}\boldsymbol{I}_p & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq^2}\\ \boldsymbol{D}_1 & \boldsymbol{I}_{pq} & \boldsymbol{O}_{pq} & \boldsymbol{D}_2 & \boldsymbol{O}_{pq,pq^2}\\ -\boldsymbol{D}_1 & \boldsymbol{O}_{pq} & \boldsymbol{I}_{pq} & \boldsymbol{D}_3 & \boldsymbol{O}_{pq,pq^2}\\ \boldsymbol{O}_{pq^2,p} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{D}_4 & \boldsymbol{I}_{pq^2}\end{bmatrix}$$

这里的 $\boldsymbol{D}_1$、$\boldsymbol{D}_2$、$\boldsymbol{D}_3$ 由式（2.9）给出。vec(・)表示括号内的矩阵按列拉直得到的列向量。

$$\boldsymbol{D}_4=-\begin{bmatrix}[(\boldsymbol{I}_q\otimes\boldsymbol{\mu}_1)+(\boldsymbol{\mu}_1\otimes\boldsymbol{I}_q)] & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & [(\boldsymbol{I}_q\otimes\boldsymbol{\mu}_2)+(\boldsymbol{\mu}_2\otimes\boldsymbol{I}_q)] & \cdots & \mathbf{0}\\ \vdots & \cdots & \cdots & \vdots\\ \mathbf{0} & \cdots & \mathbf{0} & [(\boldsymbol{I}_q\otimes\boldsymbol{\mu}_p)+(\boldsymbol{\mu}_p\otimes\boldsymbol{I}_q)]\end{bmatrix}$$

$$\boldsymbol{R}_2^*=\begin{bmatrix}\boldsymbol{E}_1 & \boldsymbol{E}_2 & \boldsymbol{O}_{pq} & \boldsymbol{O}_{pq,pq^2}\\ \boldsymbol{E}_3 & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{E}_4 & \boldsymbol{O}_{pq,pq^2}\\ \boldsymbol{O}_{pq^2,q} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{I}_{pq^2}\end{bmatrix}$$

这里的 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$、$\boldsymbol{E}_3$、$\boldsymbol{E}_4$ 由式（2.10）给出。

$$\boldsymbol{R}_3^*=\begin{bmatrix}\boldsymbol{I}_{pq} & -\boldsymbol{I}_{pq} & \boldsymbol{O}_{pq,pq^2}\\ \boldsymbol{O}_{pq^2,pq} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{I}_{pq^2}\end{bmatrix}$$

$$\boldsymbol{R}_4^*=\begin{bmatrix}\boldsymbol{I}_{pq} & \boldsymbol{O}_{pq,q^2} & \cdots & \boldsymbol{O}_{pq,q^2}\\ \boldsymbol{O}_{q^2,pq} & -(\varSigma_1^{-1}\otimes\varSigma_1^{-1}) & \cdots & \boldsymbol{O}_{q^2}\\ & \cdots & \cdots & \\ \boldsymbol{O}_{q^2,pq} & \cdots & \boldsymbol{O}_{q^2} & -(\varSigma_p^{-1}\otimes\varSigma_p^{-1})\end{bmatrix}$$

$$\boldsymbol{R}_5^* = \left[\boldsymbol{\Sigma}_1^{-1}, \cdots, \boldsymbol{\Sigma}_p^{-1}, (\boldsymbol{\mu}_{11}^{\mathrm{T}} - \boldsymbol{\mu}_{10}^{\mathrm{T}}) \otimes \boldsymbol{I}_q, \cdots, (\boldsymbol{\mu}_{p1}^{\mathrm{T}} - \boldsymbol{\mu}_{p0}^{\mathrm{T}}) \otimes \boldsymbol{I}_q \right]。$$

证明：令

$$\begin{aligned} \boldsymbol{U}_1^{*(i)} = & \left[y_1^{(i)}, \cdots, y_p^{(i)}, \boldsymbol{X}_1^{(i)\mathrm{T}} y_1^{(i)}, \cdots, \boldsymbol{X}_p^{(i)\mathrm{T}} y_p^{(i)}, \boldsymbol{X}_1^{(i)\mathrm{T}} (1 - y_1^{(i)}), \cdots, \boldsymbol{X}_p^{(i)\mathrm{T}} (1 - y_p^{(i)}), \right. \\ & \left. \boldsymbol{X}_1^{(i)\mathrm{T}}, \cdots, \boldsymbol{X}_p^{(i)\mathrm{T}}, \operatorname{vec}(\boldsymbol{X}_1^{(i)} \boldsymbol{X}_1^{(i)\mathrm{T}})^{\mathrm{T}}, \cdots, \operatorname{vec}(\boldsymbol{X}_p^{(i)} \boldsymbol{X}_p^{(i)\mathrm{T}})^{\mathrm{T}} \right]^{\mathrm{T}} \end{aligned} \tag{2.15}$$

$$\overline{\boldsymbol{U}^*}_1 = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_1^{*(i)}$$

则 $\boldsymbol{U}_1^{*(i)}$ $(i = 1, 2, \cdots, n)$ 独立同分布，期望为 $E(\boldsymbol{U}^*_1)$。

$$\begin{aligned} E(\boldsymbol{U}^*_1) = & \left[p_1, \cdots, p_p, p_1(\boldsymbol{\mu}_{11}^{\mathrm{T}} + \boldsymbol{\mu}_1^{\mathrm{T}}), \cdots, p_p(\boldsymbol{\mu}_{p1}^{\mathrm{T}} + \boldsymbol{\mu}_p^{\mathrm{T}}), \right. \\ & (1 - p_1)(\boldsymbol{\mu}_{10}^{\mathrm{T}} + \boldsymbol{\mu}_1^{\mathrm{T}}), \cdots, (1 - p_p)(\boldsymbol{\mu}_{p0}^{\mathrm{T}} + \boldsymbol{\mu}_p^{\mathrm{T}}), \\ & \left. \boldsymbol{\mu}_1^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_p^{\mathrm{T}}, \operatorname{vec}(\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^{\mathrm{T}})^{\mathrm{T}}, \cdots, \operatorname{vec}(\boldsymbol{\Sigma}_p + \boldsymbol{\mu}_p \boldsymbol{\mu}_p^{\mathrm{T}})^{\mathrm{T}} \right]^{\mathrm{T}} \end{aligned}$$

应用中心有限定理可得

$$\sqrt{n}[\overline{\boldsymbol{U}^*}_1 - E(\boldsymbol{U}_1^*)] \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega}_1^*)$$

其中

$$\begin{aligned} \boldsymbol{\Omega}_1^* = \operatorname{Cov}[& y_1, \cdots, y_p, \boldsymbol{X}_1^{\mathrm{T}} y_1, \cdots, \boldsymbol{X}_p^{\mathrm{T}} y_p, \boldsymbol{X}_1^{\mathrm{T}} (1 - y_1), \cdots, \boldsymbol{X}_p^{\mathrm{T}} (1 - y_p), \\ & \boldsymbol{X}_1^{\mathrm{T}}, \cdots, \boldsymbol{X}_p^{\mathrm{T}}, \operatorname{vec}(\boldsymbol{X}_1 \boldsymbol{X}_1^{\mathrm{T}})^{\mathrm{T}}, \cdots, \operatorname{vec}(\boldsymbol{X}_p \boldsymbol{X}_p^{\mathrm{T}})^{\mathrm{T}}]^{\mathrm{T}} \end{aligned}$$

下面考虑函数 $\boldsymbol{f}_1$：

$$\begin{array}{ccc} \mathbf{R}^{p+pq+pq+pq^2} & \to & \mathbf{R}^{p+pq+pq+pq^2} \\ \boldsymbol{u} = (\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}}, & & \left[\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}_1^{\mathrm{T}} - a_1 \boldsymbol{d}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}} - a_p \boldsymbol{d}_p^{\mathrm{T}}, \right. \\ \boldsymbol{c}_1^{\mathrm{T}}, \cdots, \boldsymbol{c}_p^{\mathrm{T}}, & \mapsto & \boldsymbol{c}_1^{\mathrm{T}} - (1 - a_1) \boldsymbol{d}_1^{\mathrm{T}}, \cdots, \boldsymbol{c}_p^{\mathrm{T}} - (1 - a_p) \boldsymbol{d}_p^{\mathrm{T}}, \\ \boldsymbol{d}_1^{\mathrm{T}}, \cdots, \boldsymbol{d}_p^{\mathrm{T}}, \boldsymbol{e}_1^{\mathrm{T}}, \cdots, \boldsymbol{e}_p^{\mathrm{T}})^{\mathrm{T}} & & \left. \boldsymbol{e}_1^{\mathrm{T}} - \operatorname{vec}(\boldsymbol{d}_1 \boldsymbol{d}_1^{\mathrm{T}})^{\mathrm{T}}, \cdots, \boldsymbol{e}_p^{\mathrm{T}} - \operatorname{vec}(\boldsymbol{d}_1 \boldsymbol{d}_1^{\mathrm{T}})^{\mathrm{T}} \right]^{\mathrm{T}} \end{array}$$

其中，$\boldsymbol{a} = (a_1, \cdots, a_p)^{\mathrm{T}} \in \mathbf{R}^p, \boldsymbol{b}_j, \boldsymbol{c}_j, \boldsymbol{d}_j \in \mathbf{R}^q, \boldsymbol{e}_j \in \mathbf{R}^{q^2}$。

令 $\boldsymbol{U}_2^* = \boldsymbol{g}_1(\overline{\boldsymbol{U}^*}_1)$，则有

$$\begin{aligned} E(\boldsymbol{U}_2^*) &= \boldsymbol{f}_1[E(\boldsymbol{U}_1^*)] \\ &= \left[p_1, \cdots, p_p, p_1 \boldsymbol{\mu}_{11}^{\mathrm{T}}, \cdots, p_p \boldsymbol{\mu}_{p1}^{\mathrm{T}}, (1 - p_1) \boldsymbol{\mu}_{10}^{\mathrm{T}}, \cdots, (1 - p_p) \boldsymbol{\mu}_{p0}^{\mathrm{T}}, \right. \\ & \quad \left. \operatorname{vec}(\boldsymbol{\Sigma}_1)^{\mathrm{T}}, \cdots, \operatorname{vec}(\boldsymbol{\Sigma}_p)^{\mathrm{T}} \right]^{\mathrm{T}} \end{aligned}$$

利用 δ-方法可得

$$\sqrt{n}[\boldsymbol{U}_2^* - E(\boldsymbol{U}_2^*)] \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Omega}_2^*)$$

其中，$\boldsymbol{\Omega}_2^* = \boldsymbol{R}_1^* \boldsymbol{\Omega}_1^* \boldsymbol{R}_1^{*\mathrm{T}}$。

$$\boldsymbol{R}_1^* = \begin{bmatrix} \boldsymbol{I}_p & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq} & \boldsymbol{O}_{p,pq^2} \\ \boldsymbol{D}_1 & \boldsymbol{I}_{pq} & \boldsymbol{O}_{pq} & \boldsymbol{D}_2 & \boldsymbol{O}_{pq,pq^2} \\ -\boldsymbol{D}_1 & \boldsymbol{O}_{pq} & \boldsymbol{I}_{pq} & \boldsymbol{D}_3 & \boldsymbol{O}_{pq,pq^2} \\ \boldsymbol{O}_{pq^2,p} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{D}_4 & \boldsymbol{I}_{pq^2} \end{bmatrix}$$

这里的 $\boldsymbol{D}_1$、$\boldsymbol{D}_2$、$\boldsymbol{D}_3$ 由式（2.9）确定，而

$$\boldsymbol{D}_4 = -\begin{bmatrix} [(\boldsymbol{I}_q \otimes \boldsymbol{\mu}_1) + (\boldsymbol{\mu}_1 \otimes \boldsymbol{I}_q)] & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & [(\boldsymbol{I}_q \otimes \boldsymbol{\mu}_2) + (\boldsymbol{\mu}_2 \otimes \boldsymbol{I}_q)] & \cdots & \boldsymbol{0} \\ \vdots & \cdots & \cdots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & [(\boldsymbol{I}_q \otimes \boldsymbol{\mu}_p) + (\boldsymbol{\mu}_p \otimes \boldsymbol{I}_q)] \end{bmatrix}$$

定义函数 $\boldsymbol{f}_2$：

$$\begin{array}{ccc} \mathbf{R}^{p+pq+pq+pq^2} & \rightarrow & \mathbf{R}^{pq+pq+pq^2} \\ \boldsymbol{u} = (\boldsymbol{a}^{\mathrm{T}}, \boldsymbol{b}_1^{\mathrm{T}}, \cdots, \boldsymbol{b}_p^{\mathrm{T}}, & \mapsto & \left(\dfrac{\boldsymbol{b}_1^{\mathrm{T}}}{a_1}, \cdots, \dfrac{\boldsymbol{b}_p^{\mathrm{T}}}{a_p}, \dfrac{\boldsymbol{c}_1^{\mathrm{T}}}{1-a_1}, \cdots, \dfrac{\boldsymbol{c}_p^{\mathrm{T}}}{1-a_p}, \right. \\ \boldsymbol{c}_1^{\mathrm{T}}, \cdots, \boldsymbol{c}_p^{\mathrm{T}}, \boldsymbol{e}_1^{\mathrm{T}}, \cdots, \boldsymbol{e}_p^{\mathrm{T}})^{\mathrm{T}} & & \left. e_1^{\mathrm{T}}, \cdots, e_p^{\mathrm{T}}\right)^{\mathrm{T}} \end{array}$$

其中，$\boldsymbol{a} = (a_1, \cdots, a_p)^{\mathrm{T}} \in \mathbf{R}^p, \boldsymbol{b}_j, \boldsymbol{c}_j \in \mathbf{R}^q, \boldsymbol{e}_j \in \mathbf{R}^{q^2}$。

令 $\boldsymbol{U}_3^* = \boldsymbol{f}_2(\boldsymbol{U}_2^*)$，则

$$\begin{aligned} E(\boldsymbol{U}_3^*) &= \boldsymbol{f}_2(E(\boldsymbol{U}_2^*)) \\ &= \left(\boldsymbol{\mu}_{11}^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_{p1}^{\mathrm{T}}, \boldsymbol{\mu}_{10}^{\mathrm{T}}, \cdots, \boldsymbol{\mu}_{p0}^{\mathrm{T}}, \mathrm{vec}(\boldsymbol{\Sigma}_1)^{\mathrm{T}}, \cdots, \mathrm{vec}(\boldsymbol{\Sigma}_p)^{\mathrm{T}}\right)^{\mathrm{T}} \end{aligned}$$

可以得到

$$\sqrt{n}[\boldsymbol{U}_3^* - E(\boldsymbol{U}_3^*)] \rightarrow_d N(\boldsymbol{0}, \boldsymbol{\Omega}_3^*)$$

其中，$\boldsymbol{\Omega}_3^* = \boldsymbol{R}_2^* \boldsymbol{\Omega}_2^* \boldsymbol{R}_2^{*\mathrm{T}}$。

$$\boldsymbol{R}_2^* = \begin{bmatrix} \boldsymbol{E}_1 & \boldsymbol{E}_2 & \boldsymbol{O}_{pq} & \boldsymbol{O}_{pq,pq^2} \\ \boldsymbol{E}_3 & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{E}_4 & \boldsymbol{O}_{pq,pq^2} \\ \boldsymbol{O}_{pq^2,q} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{O}_{pq^2,pq} & \boldsymbol{I}_{pq^2} \end{bmatrix}$$

这里的 $\boldsymbol{E}_1$、$\boldsymbol{E}_2$、$\boldsymbol{E}_3$、$\boldsymbol{E}_4$ 由式（2.10）给出。

下面定义函数 $\boldsymbol{f}_3$：

$$\begin{array}{rcl} \mathbf{R}^{pq+pq+pq^2} & \to & \mathbf{R}^{pq+pq^2} \\ \boldsymbol{u}=(\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\boldsymbol{c}_1^{\mathrm{T}},\cdots,\boldsymbol{c}_p^{\mathrm{T}},\boldsymbol{e}_1^{\mathrm{T}},\cdots,\boldsymbol{e}_p^{\mathrm{T}})^{\mathrm{T}} & \mapsto & (\boldsymbol{b}_1^{\mathrm{T}}-\boldsymbol{c}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}}-\boldsymbol{c}_p^{\mathrm{T}},\boldsymbol{e}_1^{\mathrm{T}},\cdots,\boldsymbol{e}_p^{\mathrm{T}})^{\mathrm{T}} \end{array}$$

其中，$\boldsymbol{b}_j,\boldsymbol{c}_j\in\mathbf{R}^q,\boldsymbol{e}_j\in\mathbf{R}^{q^2}$。

令

$$\begin{aligned} \boldsymbol{U}_4^* &= \boldsymbol{f}_3(\boldsymbol{U}_3^*) \\ &= \left(\hat{\boldsymbol{\mu}}_{11}^{\mathrm{T}}-\hat{\boldsymbol{\mu}}_{10}^{\mathrm{T}},\cdots,\hat{\boldsymbol{\mu}}_{p1}^{\mathrm{T}}-\hat{\boldsymbol{\mu}}_{p0}^{\mathrm{T}},\mathrm{vec}(\hat{\boldsymbol{\Sigma}}_1)^{\mathrm{T}},\cdots,\mathrm{vec}(\hat{\boldsymbol{\Sigma}}_p)^{\mathrm{T}}\right)^{\mathrm{T}} \end{aligned}$$

则有

$$\begin{aligned} E(\boldsymbol{U}_4^*) &= \boldsymbol{f}_3[E(\boldsymbol{U}_3^*)] \\ &= [\boldsymbol{\mu}_{11}^{\mathrm{T}}-\boldsymbol{\mu}_{(10)}^{\mathrm{T}},\cdots,\boldsymbol{\mu}_{p1}^{\mathrm{T}}-\boldsymbol{\mu}_{p0}^{\mathrm{T}},\mathrm{vec}(\boldsymbol{\Sigma}_1)^{\mathrm{T}},\cdots,\mathrm{vec}(\boldsymbol{\Sigma}_p)^{\mathrm{T}}]^{\mathrm{T}} \end{aligned}$$

与

$$\sqrt{n}[\boldsymbol{U}_4^*-E(\boldsymbol{U}_4^*)]\xrightarrow{d}N(\mathbf{0},\boldsymbol{\Omega}_4^*)$$

其中，$\boldsymbol{\Omega}_4^*=\boldsymbol{R}_3^*\boldsymbol{\Omega}_3^*\boldsymbol{R}_3^{*\mathrm{T}}$。

$$\boldsymbol{R}_3^*=\begin{bmatrix} \boldsymbol{I}_{pq} & -\boldsymbol{I}_{pq} & \boldsymbol{O}_{pq,\,pq^2} \\ \boldsymbol{O}_{pq^2,\,pq} & \boldsymbol{O}_{pq^2,\,pq} & \boldsymbol{I}_{pq^2} \end{bmatrix}$$

现在利用一阶近似

$$\hat{\boldsymbol{\Sigma}}_j^{-1}\doteq\boldsymbol{\Sigma}_j^{-1}-\boldsymbol{\Sigma}_j^{-1}(\hat{\boldsymbol{\Sigma}}_j-\boldsymbol{\Sigma}_j)\boldsymbol{\Sigma}_j^{-1}$$

得到

$$\sqrt{n}\left(\begin{bmatrix}\hat{\boldsymbol{\mu}}_{11}-\hat{\boldsymbol{\mu}}_{10}\\ \vdots \\ \hat{\boldsymbol{\mu}}_{p1}-\hat{\boldsymbol{\mu}}_{p0}\\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_1^{-1})\\ \vdots \\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_p^{-1})\end{bmatrix}-\begin{bmatrix}\boldsymbol{\mu}_{11}-\boldsymbol{\mu}_{10}\\ \vdots \\ \boldsymbol{\mu}_{p1}-\boldsymbol{\mu}_{p0}\\ \mathrm{vec}(\boldsymbol{\Sigma}_1^{-1})\\ \vdots \\ \mathrm{vec}(\boldsymbol{\Sigma}_p^{-1})\end{bmatrix}\right)\doteq\boldsymbol{R}_4^*\sqrt{n}\left(\begin{bmatrix}\hat{\boldsymbol{\mu}}_{11}-\hat{\boldsymbol{\mu}}_{10}\\ \vdots \\ \hat{\boldsymbol{\mu}}_{p1}-\hat{\boldsymbol{\mu}}_{p0}\\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_1)\\ \vdots \\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_p)\end{bmatrix}-\begin{bmatrix}\boldsymbol{\mu}_{11}-\boldsymbol{\mu}_{10}\\ \vdots \\ \boldsymbol{\mu}_{p1}-\boldsymbol{\mu}_{p0}\\ \mathrm{vec}(\boldsymbol{\Sigma}_1)\\ \vdots \\ \mathrm{vec}(\boldsymbol{\Sigma}_p)\end{bmatrix}\right)$$

其中

$$\boldsymbol{R}_4^*=\begin{bmatrix} \boldsymbol{I}_{pq} & \boldsymbol{O}_{pq,\,q^2} & \cdots & \boldsymbol{O}_{pq,\,q^2} \\ \boldsymbol{O}_{q^2,\,pq} & -(\boldsymbol{\Sigma}_1^{-1}\otimes\boldsymbol{\Sigma}_1^{-1}) & \cdots & \boldsymbol{O}_{q^2} \\ & \cdots & \cdots & \\ \boldsymbol{O}_{q^2,\,pq} & \cdots & \boldsymbol{O}_{q^2} & -(\boldsymbol{\Sigma}_p^{-1}\otimes\boldsymbol{\Sigma}_p^{-1}) \end{bmatrix}$$

因此，有

$$\sqrt{n}\left(\begin{bmatrix}\hat{\mu}_{11}-\hat{\mu}_{10}\\ \vdots \\ \hat{\mu}_{p1}-\hat{\mu}_{p0}\\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_1^{-1})\\ \vdots \\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_p^{-1})\end{bmatrix}-\begin{bmatrix}\mu_{11}-\mu_{10}\\ \vdots \\ \mu_{p1}-\mu_{p0}\\ \mathrm{vec}(\boldsymbol{\Sigma}_1^{-1})\\ \vdots \\ \mathrm{vec}(\boldsymbol{\Sigma}_p^{-1})\end{bmatrix}\right)\xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Omega}_4^{*})$$

其中

$$\boldsymbol{\Omega}_4^{*}=\boldsymbol{R}_4^{*}\boldsymbol{\Omega}_4^{*}\boldsymbol{R}_4^{*\mathrm{T}}$$

最后定义函数 $\boldsymbol{f}_4$：

$$\begin{aligned} \mathbf{R}^{pq+pq^2} &\to \mathbf{R}^q \\ \boldsymbol{u}=[\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\mathrm{vec}(\boldsymbol{C}_1)^{\mathrm{T}},\cdots,\mathrm{vec}(\boldsymbol{C}_p)^{\mathrm{T}}]^{\mathrm{T}} &\mapsto \sum_{j=1}^{p}\boldsymbol{C}_j\boldsymbol{b}_j \end{aligned}$$

其中，$\boldsymbol{b}_j\in\mathbf{R}^q,\ \boldsymbol{C}_j\in\mathbf{R}^{q\times q}(j=1,2,\cdots,p)$，则有

$$\hat{\gamma}=\frac{1}{p}\boldsymbol{f}_4[\hat{\mu}_{11}^{\mathrm{T}}-\hat{\mu}_{10}^{\mathrm{T}},\cdots,\hat{\mu}_{p1}^{\mathrm{T}}-\hat{\mu}_{p0}^{\mathrm{T}},\mathrm{vec}(\hat{\boldsymbol{\Sigma}}_1^{-1})^{\mathrm{T}},\cdots,\mathrm{vec}(\hat{\boldsymbol{\Sigma}}_p^{-1})^{\mathrm{T}}]^{\mathrm{T}}$$

$$\gamma=\frac{1}{p}\boldsymbol{f}_4[\mu_{11}^{\mathrm{T}}-\mu_{10}^{\mathrm{T}},\cdots,\mu_{p1}^{\mathrm{T}}-\mu_{p0}^{\mathrm{T}},\mathrm{vec}(\boldsymbol{\Sigma}_1^{-1})^{\mathrm{T}},\cdots,\mathrm{vec}(\boldsymbol{\Sigma}_p^{-1})^{\mathrm{T}}]^{\mathrm{T}}$$

由 δ-方法，得

$$\sqrt{n}(\hat{\gamma}-\gamma)\xrightarrow{d} N(\mathbf{0},\ \boldsymbol{\Omega}_{\gamma}^{*})$$

其中

$$\boldsymbol{\Omega}_{\gamma}^{*}=\frac{1}{p^2}\boldsymbol{R}_5^{*}\boldsymbol{\Omega}_5^{*}\boldsymbol{R}_5^{*\mathrm{T}}$$

$$\boldsymbol{R}_5^{*}=\left[\boldsymbol{\Sigma}_1^{-1},\cdots,\boldsymbol{\Sigma}_p^{-1},(\mu_{11}^{\mathrm{T}}-\mu_{10}^{\mathrm{T}})\otimes\boldsymbol{I}_q,\cdots,(\mu_{p1}^{\mathrm{T}}-\mu_{p0}^{\mathrm{T}})\otimes\boldsymbol{I}_q\right]$$

证毕。

同前面的讨论一样，在实际应用中用 $\hat{\boldsymbol{p}}_j$、$\hat{\mu}_{j1}$、$\hat{\mu}_{j0}$、$\hat{\mu}_j$ 和 $\hat{\boldsymbol{\Sigma}}_j$ 代替 $\boldsymbol{p}_j$、μ_{j1}、μ_{j0}、μ_j 和 $\boldsymbol{\Sigma}_j$，便可以得到 $\boldsymbol{R}_1^{*},\cdots,\boldsymbol{R}_5^{*}$ 的相合估计。进一步，可以得到渐近协方差阵 $\boldsymbol{\Omega}_{\gamma}^{*}$ 的一个相合估计。

2.4 假设检验

2.3 节中我们得到了估计的渐近正态性，并且可以得到渐近协方差阵的相合估计，在这一节中我们利用这些结果讨论回归系数 $\boldsymbol{\beta}$ 的假设检验。在这一节中为了方便起见，不论是 $\boldsymbol{\Sigma}_j$ 已知还是未知，一律用 $\hat{\boldsymbol{\Omega}}_\gamma$ 表示渐近协方差阵的相合估计。

1．检验 β 的线性假设

线性假设的原假设为

$$H_0: \quad \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=\boldsymbol{0} \tag{2.16}$$

其中，$\boldsymbol{L}$ 为一个 $q\times r\ (r\leqslant q)$ 列满秩矩阵。构造检验统计量为

$$T=n\hat{\gamma}^{\mathrm{T}}\boldsymbol{L}(\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_\gamma\boldsymbol{L})^{-1}\boldsymbol{L}^{\mathrm{T}}\hat{\gamma} \tag{2.17}$$

定理 2.4.1 在定理 2.3.1 或定理 2.3.1 的条件下，如果原假设式（2.16）成立，则检验统计量式（2.17）的渐近分布为

$$T\xrightarrow{d}\chi_r^2$$

证明：如果原假设式（2.16）成立，则

$$\boldsymbol{L}^{\mathrm{T}}\gamma=\boldsymbol{0}$$

因此可得

$$T=n\hat{\gamma}^{\mathrm{T}}\boldsymbol{L}(\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_\gamma\boldsymbol{L})^{-1}\boldsymbol{L}^{\mathrm{T}}\hat{\gamma}=n(\hat{\gamma}^{\mathrm{T}}-\gamma^{\mathrm{T}})\boldsymbol{L}(\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_\gamma\boldsymbol{L})^{-1}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)$$

由定理 2.3.1 与定理 2.3.2，得到

$$\sqrt{n}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Omega}_\gamma\boldsymbol{L})$$

因为 $\boldsymbol{L}$ 是一个列满秩矩阵，所以 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Omega}_\gamma\boldsymbol{L}$ 是一个 $r\times r$ 正定矩阵。因此，有

$$T\xrightarrow{d}\chi_r^2$$

其中，χ_r^2 是自由度为 r 的卡方分布，证毕。

应用定理 2.4.1，可以得到检验式（2.16）的显著性水平为 α 的拒绝域为

$$\left\{(\boldsymbol{Y},\boldsymbol{X}): T\geqslant\chi_{r,1-\alpha}^2\right\}$$

式中，$\chi_{r,1-\alpha}^2$ 是自由度为 r 的卡方分布下 $1-\alpha$ 分位数。

2．回归变量的选择

原假设为

$$H_0: \quad \beta_h=0$$

如果原假设成立，则变量 $\boldsymbol{X}_j\,(j=1,2,\cdots,p)$ 的第 h 个分量对模型的贡献很小，这样就可以将各 $\boldsymbol{X}_j$ 中的第 h 个分量和 $\boldsymbol{\beta}$ 中的第 h 个分量删去从而使模型简化。

检验统计量仍然由式（2.17）给出，其中 $\boldsymbol{L}=(0,\cdots,0,1,0,\cdots,0)^{\mathrm{T}}$，即令 $\boldsymbol{L}$ 中的第 h 个分量为 1，其他分量均为 0，则这个检验的显著性水平为 α 的拒绝域为

$$\left\{(\boldsymbol{Y},\boldsymbol{X}),\ T\geqslant\chi^2_{1,1-\alpha}\right\}$$

式中，$\chi^2_{1,1-\alpha}$ 是自由度为 1 的卡方分布下 $1-\alpha$ 分位数。

3．单边假设的检验

因为 $\boldsymbol{\beta}$ 可以被估计到只相差一个正常数乘积，所以可以进行单边假设的检验：

$$H_{10}\!:\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}\leqslant 0\ \leftrightarrow\ H_{11}\!:\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}>0 \tag{2.18}$$

或

$$H_{20}\!:\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}\geqslant 0\ \leftrightarrow\ H_{21}:\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}<0 \tag{2.19}$$

式中，$\boldsymbol{L}$ 是一个 $p\times 1$ 非零向量。令检验统计量为

$$T=\frac{\sqrt{n}\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\gamma}}}{\sqrt{\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}\boldsymbol{L}}} \tag{2.20}$$

检验统计量 T 的渐近分布为 $N\left(\dfrac{\sqrt{n}\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\gamma}}{\sqrt{\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\Omega}\boldsymbol{L}}},1\right)$，则检验式（2.18）和式（2.19）的显著性水平为 α 的拒绝域为

$$\{(\boldsymbol{Y},\boldsymbol{X})\!:\ T>z_{1-\alpha}\}$$
$$\{(\boldsymbol{Y},\boldsymbol{X})\!:\ T<z_{\alpha}\}$$

式中，z_α 和 $z_{1-\alpha}$ 分别是标准正态分布 $N(0,1)$ 的 α 和 $1-\alpha$ 的下分位数。

2.5　模拟研究

2.5.1　点估计的模拟研究

令 $\boldsymbol{X}$ 为一个 5×6 随机矩阵，$\mathrm{vec}(\boldsymbol{X})\sim N_{30}(\boldsymbol{\mu},\boldsymbol{\Sigma})$，扰动项 $\boldsymbol{\varepsilon}\sim N_5(\boldsymbol{0},\boldsymbol{\Sigma}_{\varepsilon})$。其中 $\boldsymbol{\mu}$=[1.9，−0.9，1.7，1.8，−0.9，2.7，1，−5，2，0.9，−4.5，2，2.7，−1.1，0.5，1.8，−1.8，1.4，−0.9，1.1，−1.4，−0.9，1.8，1.8，−2.3，1.4，−1.6，−1.8，1.8，3.2]$^{\mathrm{T}}$。

$$\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Sigma}_2$$

$$\boldsymbol{\Sigma}_1=\begin{bmatrix}1 & 0.5 & -0.5 & 0.2 & -0.3\\ 0.5 & 2 & 0.3 & -0.2 & 0.1\\ -0.5 & 0.3 & 1 & 0.2 & -0.2\\ 0.2 & -0.2 & 0.2 & 2 & 0.2\\ -0.3 & 0.1 & -0.2 & 0.2 & 1\end{bmatrix}$$

$$\boldsymbol{\Sigma}_2=\begin{bmatrix}2 & -0.4 & 0.3 & 0.5 & -0.2 & 0.1\\ -0.4 & 1 & -0.2 & 0.3 & 0.2 & -0.3\\ 0.3 & -0.2 & 2 & -0.3 & 0.1 & -0.2\\ 0.5 & 0.3 & -0.3 & 1 & 0.2 & -0.2\\ -0.2 & 0.2 & 0.1 & 0.2 & 2 & 0.1\\ 0.1 & -0.3 & -0.2 & -0.2 & 0.1 & 1\end{bmatrix}$$

$$\boldsymbol{\Sigma}_\varepsilon=\begin{bmatrix}1 & 0.1 & 0.2 & -0.1 & -0.2\\ 0.1 & 1.5 & -0.3 & 0.1 & 0.2\\ 0.2 & -0.3 & 2 & -0.4 & -0.1\\ -0.1 & 0.1 & -0.4 & 2 & 0.2\\ -0.2 & 0.2 & -0.1 & 0.2 & 1\end{bmatrix}$$

回归系数 $\boldsymbol{\beta}$ 的真值为 $(1,2,3,-2,-1,0)^{\mathrm{T}}$ 。

由模型产生样本容量为 n=50，100，300，500 的随机数，重复 500 次估计回归系数方向的均值及其标准差。为了便于比较，在这里列出 $\hat{\gamma}_2/\hat{\gamma}_1,\cdots,\hat{\gamma}_6/\hat{\gamma}_1$ 的值。

评价方向估计的另一个标准是：

$$c=\cos(\hat{\gamma},\boldsymbol{\beta})=\frac{\hat{\gamma}^{\mathrm{T}}\boldsymbol{\beta}}{\sqrt{(\hat{\gamma}^{\mathrm{T}}\hat{\gamma})}\sqrt{(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})}}$$

这里我们也计算 c 值。模拟结果列入表 2.1、表 2.2 和表 2.3 中。

表 2.1　当 Σ_j 已知时回归系数估计的均值与标准差（一）（括号内为标准差）

n	模拟结果	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
50	均值	0.4007	0.7969	1.2007	−0.7967	−0.3990	0.0048
	标准差	（0.1204）	（0.1047）	（0.1084）	（0.1104）	（0.1152）	（0.1073）
100	均值	0.3708	0.7532	1.1280	−0.7535	−0.3730	−0.0006
	标准差	（0.0627）	（0.0616）	（0.0614）	（0.0647）	（0.0669）	（0.0548）
300	均值	0.3736	0.7454	1.1156	−0.7440	−0.3712	−0.0049
	标准差	（0.0343）	（0.0321）	（0.0344）	（0.0355）	（0.0334）	（0.0362）
500	均值	0.3708	0.7384	1.1066	−0.7397	−0.3672	−0.0018
	标准差	（0.0426）	（0.0411）	（0.0354）	（0.0401）	（0.0437）	（0.0428）

表 2.2　$\hat{\gamma}_j/\hat{\gamma}_1$ 的均值和标准差（一）（括号内为标准差）

n	模拟结果	$\hat{\gamma}_2/\hat{\gamma}_1$	$\hat{\gamma}_3/\hat{\gamma}_1$	$\hat{\gamma}_4/\hat{\gamma}_1$	$\hat{\gamma}_5/\hat{\gamma}_1$	$\hat{\gamma}_6/\hat{\gamma}_1$
50	均值	2.2692	3.4241	−2.2490	−1.1221	−0.0240
	标准差	（0.776）	（1.3955）	（0.7064）	（0.4341）	（0.3172）
100	均值	2.1026	3.1488	−2.0997	−1.0338	−0.0002
	标准差	（0.3531）	（0.5893）	（0.3321）	（0.2126）	（0.1908）
300	均值	2.0145	3.0138	−2.0080	−1.0016	−0.0123
	标准差	（0.1773）	（0.2701）	（0.1532）	（0.1043）	（0.0979）
500	均值	2.0157	3.0259	−2.0143	−1.0090	−0.0008
	标准差	（0.1318）	（0.2143）	（0.1200）	（0.0841）	（0.0761）

表 2.3　c 值的均值与标准差（一）（括号内为标准差）

n	50	100	300	500
c	0.9897	0.9960	0.9982	0.9993
标准差	（0.0065）	（0.0027）	（0.0008）	（0.0005）

为了评价本章给出的方法，我们与模拟极大似然方法比较。这里利用 GHK 模拟器进行模拟，GHK 模拟器在文献[16]中有详细介绍。GHK 模拟方法中的参数是连续平滑的，并且在模拟估计方法中有较快的运算速度，计算程序公布在 Vassilis A. Hajivassiliou 的网页上，网址：http:// econ.lse.ac.uk/vassilis。GHK 模拟过程计算高维积分需要从多元正态分布中抽取随机数，每次积分时抽取的模拟随机数数量 R 与文献[16]中的相同，取 R=5000。我们把模拟结果列入表 2.4、表 2.5 和表 2.6 中。

表 2.4　当 Σ_j 未知时回归系数估计的均值与标准差（二）（括号内是标准差）

n	模拟结果	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
50	均值	1.2853	2.7802	4.0242	−2.6980	−1.2655	−0.0204
	标准差	（0.2707）	（0.5055）	（0.7784）	（0.5178）	（0.2384）	（0.2342）
100	均值	0.9757	1.8847	2.9367	−1.9327	−1.0393	−0.0571
	标准差	（0.1364）	（0.2925）	（0.3449）	（0.2043）	（0.1450）	（0.0983）
300	均值	1.0218	2.0505	3.0719	−2.0588	−1.0245	−0.0050
	标准差	（0.1057）	（0.1418）	（0.2219）	（0.1392）	（0.1004）	（0.0852）
500	均值	1.0125	2.0330	3.0379	−2.0286	−1.0134	−0.0012
	标准差	（0.0782）	（0.1123）	（0.1629）	（0.1117）	（0.0795）	（0.0712）

表 2.5 $\hat{\gamma}_j/\hat{\gamma}_1$ 的均值和标准差（二）（括号内为标准差）

n	模拟结果	$\hat{\gamma}_2/\hat{\gamma}_1$	$\hat{\gamma}_3/\hat{\gamma}_1$	$\hat{\gamma}_4/\hat{\gamma}_1$	$\hat{\gamma}_5/\hat{\gamma}_1$	$\hat{\gamma}_6/\hat{\gamma}_1$
50	均值	2.2332	3.2003	−2.1940	−0.9808	−0.0066
	标准差	（0.5328）	（0.5828）	（0.4387）	（0.3295）	（0.1987）
100	均值	1.9585	3.0600	−2.0146	−1.0789	−0.0675
	标准差	（0.2745）	（0.4015）	（0.2014）	（0.1401）	（0.1221）
300	均值	2.0247	3.0370	−2.0302	−1.0098	−0.0610
	标准差	（0.1659）	（0.2493）	（0.1420）	（0.0972）	（0.0849）
500	均值	2.0172	3.0143	−2.0123	−1.0046	−0.0009
	标准差	（0.1223）	（0.1872）	（0.1247）	（0.0870）	（0.0817）

表 2.6 c 值的均值与标准差（二）（括号内为标准差）

n	50	100	300	500
c	0.9921	0.9970	0.9986	0.9994
标准差	（0.0054）	（0.0018）	（0.0007）	（0.0005）

因为本章给出的估计和模拟极大似然估计相差一个乘积尺度，所以不易从表 2.1 和表 2.4 中观察这两种估计的结果。我们可以观察表 2.2 和表 2.5 及表 2.3 和表 2.6。在取小样本时，模拟极大似然估计要好于本章给出的方法，这并不奇怪，因为模拟极大似然估计是渐近有效的。在取大样本时，两种方法有相近的估计效果，这是因为本章给出的方法在大样本性质上是渐近正态的。另外，模拟扰动对模拟极大似然估计也存在影响。从模拟的结果看本章给出的方法有良好的大样本性质。

本章所述估计的另一个优点是无须对扰动项做任何假定。第二个模拟中的自变量与第一个模拟中的自变量分布相同，回归系数真值不变。扰动项 $\boldsymbol{\varepsilon}=(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)^{\mathrm{T}}$ 服从球面上的均匀分布，$\boldsymbol{\varepsilon}$ 的随机数由下面的方式产生：5 维随机向量 ξ 的每个分量都独立并且服从 $N(0,1)$ 分布，$\boldsymbol{\varepsilon}=1.5\times\dfrac{\xi}{\|\xi\|}$。其中，$\|\xi\|$ 为向量 ξ 的模。实验重复 500 次，模拟结果列入表 2.7～表 2.9 中。

表 2.7 当 Σ_j 未知时回归系数估计的均值与标准差（三）（扰动为球面分布，括号内是标准差）

n	模拟结果	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
50	均值	0.4062	0.8353	1.2489	−0.8328	−0.4190	−0.0023
	标准差	（0.1069）	（0.1124）	（0.1118）	（0.1086）	（0.1095）	（0.1103）
100	均值	0.3979	0.8057	1.2101	−0.8005	−0.4005	−0.0013

续表

n	模拟结果	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$
100	标准差	（0.0700）	（0.0725）	（0.0668）	（0.0721）	（0.0723）	（0.0716）
300	均值	0.3998	0.8003	1.1975	−0.7991	−0.3983	−0.0006
	标准差	（0.0391）	（0.0417）	（0.0411）	（0.0375）	（0.0382）	（0.0406）
500	均值	0.3986	0.7943	1.1943	−0.7965	−0.3987	−0.0002
	标准差	（0.0299）	（0.0292）	（0.0303）	（0.0307）	（0.0293）	（0.0305）

表 2.8　$\hat{\gamma}_j/\hat{\gamma}_1$ 的均值和标准差（三）（括号内为标准差）

n	模拟结果	$\hat{\gamma}_2/\hat{\gamma}_1$	$\hat{\gamma}_3/\hat{\gamma}_1$	$\hat{\gamma}_4/\hat{\gamma}_1$	$\hat{\gamma}_5/\hat{\gamma}_1$	$\hat{\gamma}_6/\hat{\gamma}_1$
50	均值	2.2468	3.3674	−2.2206	−1.1178	0.0023
	标准差	（0.6265）	（1.0765）	（0.5553）	（0.3649）	（0.3269）
100	均值	2.0997	3.0253	−2.0178	−1.0052	0.0002
	标准差	（ 0.3630）	（0.5680 ）	（0.3243）	（0.2198）	（0.1936）
300	均值	2.0226	3.0370	−2.0302	−1.0098	−0.0022
	标准差	（ 0.1895）	（0.2767）	（0.1641）	（0.1113）	（0.1026）
500	均值	2.0050	3.0146	−2.0091	−1.0054	−0.0002
	标准差	（0.1402）	（0.2198）	（0.1226）	（0.0841）	（0.0769）

表 2.9　c 值的均值与标准差（三）（括号内为标准差）

n	50	100	300	500
c	0.9910	0.9958	0.9987	0.9992
标准差	（0.0062）	（0.0027）	（0.0008）	（0.0005）

当误差项的分布为非正态分布时，本章给出的估计结果与前面的结果相似。结果说明本章给出的方法只要求误差项与自变量相互独立，而无须对误差项的分布做任何假定。

2.5.2　线性假设的检验

令 $\boldsymbol{X}$ 为一个 5×4 的随机矩阵，$\text{vec}(\boldsymbol{X})\sim N_{20}(\boldsymbol{\mu},\boldsymbol{\Sigma})$，$\boldsymbol{\varepsilon}\sim N_5(\mathbf{0},\boldsymbol{\Sigma}_\varepsilon)$，其中 $\boldsymbol{\mu}=[1.7,-2.4,0.9,-0.6,1.7,0,1,-0.4,-1,-1.6,-1.5,0.9,-1,0.2,-1.7,1,-0.6,1.8,1.8,0.5]^{\mathrm{T}}$

$$\boldsymbol{\Sigma}=\begin{bmatrix}1 & 0.2 & \cdots & 0.2\\ 0.2 & 1 & \cdots & 0.2\\ & \cdots & \cdots & \\ 0.2 & 0.2 & \cdots & 1\end{bmatrix}$$

$$\boldsymbol{\Sigma}_{\varepsilon}=\begin{bmatrix}2 & -0.2 & 0.1 & -0.1 & -0.2\\ -0.2 & 1 & -0.1 & 0.3 & 0.2\\ 0.1 & -0.1 & 3 & -0.4 & -0.2\\ -0.1 & 0.3 & -0.4 & 2 & -0.3\\ -0.2 & 0.2 & -0.2 & -0.3 & 1.5\end{bmatrix}$$

令

$$\boldsymbol{L}^{\mathrm{T}}=\begin{bmatrix}1 & 1 & 1 & 1\\ -2 & 1 & 1 & -2\end{bmatrix}$$

$$\boldsymbol{\beta}=(-1,-1,1,1)^{\mathrm{T}}$$

原假设为 H_0: $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=0$。由模型产生容量大小为 $n=50,100,300,500,1000$ 的样本并且重复 10000 次，获得检验犯第一类错误的频率，结果列入表 2.10 中。

表 2.10　检验犯第一类错误的频率

样本容量	Σ_j 已知			Σ_j 未知		
	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.025$	$\alpha=0.10$	$\alpha=0.05$	$\alpha=0.025$
n=50	0.1275	0.0698	0.0380	0.1422	0.0846	0.0502
n=100	0.1127	0.0609	0.0339	0.1296	0.0736	0.0405
n=300	0.1083	0.0532	0.0277	0.1081	0.0570	0.0283
n=500	0.1015	0.0506	0.0264	0.1032	0.0531	0.0272
n=1000	0.0991	0.0476	0.0242	0.0988	0.0465	0.0265

由模拟结果发现，犯第一类错误的频率随着样本容量的增加而接近真实的显著性水平，当样本容量为 300 时，已经很接近真实水平了。

下面利用计算机模拟来考察上述检验的势。回归系数 $\boldsymbol{\beta}$ 的真值分别取下面的不同值。$\boldsymbol{\beta}=(-1,-1.25,1,1)^{\mathrm{T}}$，$\boldsymbol{\beta}=(-1,-1,1.25,1)^{\mathrm{T}}$，$\boldsymbol{\beta}=(-1,-1.5,1,1)^{\mathrm{T}}$，$\boldsymbol{\beta}=(-1,-1,1.5,1)^{\mathrm{T}}$，$\boldsymbol{\beta}=(-1,-1,2,1)^{\mathrm{T}}$。$\boldsymbol{L}$ 的取值仍为

$$\boldsymbol{L}^{\mathrm{T}}=\begin{bmatrix}1 & 1 & 1 & 1\\ -2 & 1 & 1 & -2\end{bmatrix}$$

满足

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=\begin{bmatrix}-0.25\\ -0.25\end{bmatrix},\begin{bmatrix}0.25\\ 0.25\end{bmatrix},\begin{bmatrix}-0.5\\ -0.5\end{bmatrix},\begin{bmatrix}0.5\\ 0.5\end{bmatrix},\begin{bmatrix}1\\ 1\end{bmatrix}$$

令检验的显著性水平 $\alpha=0.05$，样本容量为 100，300，500，1000。实验重复 10000 次，计算检验的势，模拟结果如表 2.11 所示。

表 2.11　检验的势

$\alpha=0.05$, Σ_j 已知时检验的势					
$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	（−0.25，−0.25）	（0.25，0.25）	（−0.5，−0.5）	（0.5，0.5）	（1，1）
n=100	0.2332	0.2457	0.5059	0.5097	0.9128
n=300	0.4638	0.4515	0.9707	0.9844	1
n=500	0.6657	0.6738	1	1	1
n=1000	0.9385	0.9502	1	1	1
$\alpha=0.05$, Σ_j 未知时检验的势					
$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	（−0.25，−0.25）	（0.25，0.25）	（−0.5，−0.5）	（0.5，0.5）	（1，1）
n=100	0.2877	0.2806	0.5736	0.5863	0.9673
n=300	0.4925	0.5313	0.9724	0.9871	1
n=500	0.7138	0.7159	1	1	1
n=1000	0.9420	0.9511	1	1	1

下面模拟 $\boldsymbol{L}$ 取值不同时，不同检验的势。$\boldsymbol{L}^{\mathrm{T}}$ 取下面不同的值：

$$\begin{bmatrix} 1 & 1 & 1 & 1.5 \\ -2 & 1 & 1 & -2 \end{bmatrix},\ \begin{bmatrix} 1 & 1 & 1 & 1 \\ -2 & 1 & 1.5 & -2 \end{bmatrix},\ \begin{bmatrix} 1 & 1 & 1 & 1.5 \\ -2 & 1 & 1 & -2 \end{bmatrix},$$
$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -2 & 1 & 2 & -2 \end{bmatrix},\ \begin{bmatrix} 1 & 1 & 1 & 2 \\ -2 & 1 & 2 & -2 \end{bmatrix}$$

满足

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=\begin{bmatrix} 0.5 \\ 0 \end{bmatrix},\begin{bmatrix} 0 \\ 0.5 \end{bmatrix},\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix},\begin{bmatrix} 0 \\ 1 \end{bmatrix},\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

令检验的显著性水平 $\alpha=0.05$，样本容量为 100，300，500，1000。实验重复 10000 次，计算检验的势，填入表 2.12 中。

表 2.12　检验的势

$\alpha=0.05$, Σ_j 已知时检验的势					
$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	（0.5，0）	（0，0.5）	（0.5，0.5）	（0，1）	（1，1）
n=100	0.3642	0.3027	0.7119	0.9312	1
n=300	0.8093	0.7535	0.9673	1	1
n=500	0.9647	0.9228	1	1	1
n=1000	1	1	1	1	1
$\alpha=0.05$, Σ_j 未知时检验的势					
$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	（0.5，0）	（0，0.5）	（0.5，0.5）	（0，1）	（1，1）
n=100	0.4547	0.3976	0.7621	0.9568	1
n=300	0.8775	0.6723	0.9895	1	1
n=500	0.9858	0.9089	1	1	1
n=1000	1	1	1	1	1

2.5.3 回归变量的选择

令 $\boldsymbol{X}$ 和 $\boldsymbol{\varepsilon}$ 与 2.5.1 节中的分布相同，回归系数真值为 $\boldsymbol{\beta}=[1,2,0,0.2,-1,1]$，这里我们只考虑 $\boldsymbol{\Sigma}_j$ 未知的情况。首先，由模型产生一个样本容量为 n=500 的随机数进行点估计，估计的结果为

$$\hat{\gamma}=[0.5338,\ 1.0521,\ -0.0206,\ 0.0839,\ -0.5340,\ 0.5231]^{\mathrm{T}}$$

发现 $\hat{\gamma}_3=-0.0206$ 和 $\hat{\gamma}_4=0.0839$，比其他分量更接近 0。因此分别检验下面两个假设：

$$H_0:\ \beta_3=0,\quad H_0':\beta_4=0$$

分别重复样本容量为 n=300,500,1000 的 10000 实验以获得检验统计量 T 落入以下区间的频率：$T<\chi^2_{0.90},\chi^2_{0.90}\leqslant T<\chi^2_{0.95},\chi^2_{0.95}\leqslant T<\chi^2_{0.975},\chi^2_{0.975}<T$。其中，$\chi^2_\alpha$ 表示自由度为 1 的卡方分布的下 α 分位数。模拟结果列入表 2.13 中，并且表中能确定拒绝原假设显著性的大小。

表 2.13　统计量 T 落入各区间的频率

H_0	$T<\chi^2_{0.90}$	$\chi^2_{0.90}\leqslant T<\chi^2_{0.95}$	$\chi^2_{0.95}\leqslant T<\chi^2_{0.975}$	$\chi^2_{0.975}\leqslant T$
n=300	0.8879	0.0543	0.0230	0.0348
n=500	0.8962	0.0513	0.0261	0.0264
n=1000	0.9026	0.0490	0.0256	0.0228
H_0'	$T<\chi^2_{0.90}$	$\chi^2_{0.90}\leqslant T<\chi^2_{0.95}$	$\chi^2_{0.95}\leqslant T<\chi^2_{0.975}$	$\chi^2_{0.975}\leqslant T$
n=300	0.1354	0.0698	0.0876	0.7072
n=500	0.0497	0.0309	0.0411	0.8783
n=1000	0	0.0011	0.0023	0.9966

第 3 章　多元秩-序模型回归系数的估计

本章考虑多元秩-序模型回归系数的估计问题。设 p 维随机变量 $\boldsymbol{Y}^*$ 与 $q\times p$ 的矩阵 $\boldsymbol{X}$ 有如下关系：

$$\boldsymbol{Y}^* = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.1}$$

式中，$\boldsymbol{X}$ 与 $\boldsymbol{\varepsilon}$ 独立，$\boldsymbol{\beta} \in \mathbf{R}^q$，是未知的回归系数。$\boldsymbol{Y}$ 和 $\boldsymbol{Y}^*$ 之间的关系由下文定义的映射确定。

令 r_j 为向量 $\boldsymbol{Y}^*$ 中小于等于第 j 个分量的分量个数，这时称 r_j 为第 j 分量的秩。定义

$$\boldsymbol{Y} = \tau(\boldsymbol{Y}^*) = (r_1, r_2, \cdots, r_p)^{\mathrm{T}} \tag{3.2}$$

这时我们称 $\boldsymbol{Y}$ 为 $\boldsymbol{Y}^*$ 的一个秩序。

本章研究了多元秩-序模型（MRO 模型）的逆回归估计。在 3.1 节中，将引理 2.1.2 推广到多元秩-序模型中，得到了多元秩-序模型的逆回归性质。在 3.2 节中，根据多元秩-序模型的逆回归性质，当自变量满足线性条件 A1 时，提出了一种计算简单的估计方法。3.2 节的估计方法可以在不预先对误差分布做具体假定的情况下，得到对回归系数方向的估计，并且这个估计与回归系数只相差一个正常数因子。在 3.3 节中，证明了估计是 $\sqrt{n}$ 相合的。由模型的可识别性，可以直接利用这个估计进行建模和预报。3.4 节给出了估计的模拟结果。由于没有得到估计的渐近分布，3.5 节考虑利用 Bootstrap 方法对回归系数进行假设检验，给出检验统计量，证明检验统计量分布的 Bootstrap 估计是强相合的。进一步证明了 Bootstrap 检验是渐近精确的和相合的，并且给出了 Bootstrap 检验水平和势的模拟结果。

3.1　多元秩-序模型的逆回归性质

在多元秩-序模型中，对于任意两个回归方程的差有

$$y_h^* - y_j^* = (\boldsymbol{X}_h^{\mathrm{T}} - \boldsymbol{X}_j^{\mathrm{T}})\beta + \varepsilon_h - \varepsilon_j$$

有扰动项 $\varepsilon_h - \varepsilon_j$ 与自变量 $\boldsymbol{X}_h^{\mathrm{T}} - \boldsymbol{X}_j^{\mathrm{T}}$ 独立。

记

$$\boldsymbol{Z}_j = \boldsymbol{X}_j - E(\boldsymbol{X}_j), \boldsymbol{\Sigma}_{hj} = \mathrm{Cov}(\boldsymbol{X}_h - \boldsymbol{X}_j), h \neq j, h、j = 1, 2, \cdots, p$$

并且设 $\boldsymbol{\Sigma}_{hj} > 0$，可以得到引理 3.1.1。

引理 3.1.1 在多元秩-序模型中，若 $\boldsymbol{X}_h, \boldsymbol{X}_j$ $(h、j = 1, 2, \cdots, p)$ 满足线性条件 A1，则有

$$E(\boldsymbol{Z}_h - \boldsymbol{Z}_j | y_h^* - y_j^*) \in \boldsymbol{S}(\boldsymbol{\Sigma}_{hj}\boldsymbol{\beta}), \ h \neq j, h、j = 1, \cdots, p \tag{3.3}$$

其中，$\boldsymbol{S}(\boldsymbol{\Sigma}_{hj}\boldsymbol{\beta})$ 表示由列向量 $\boldsymbol{\Sigma}_{hj}\boldsymbol{\beta}$ 所张成的线性空间。

由引理 2.1.2 可以直接得到引理 3.1.2。

引理 3.1.2 设 W_j 与 δ_j 独立 $(j = 1,2)$，令

$$V_j = W_j + \delta_j, \ j = 1, 2 \tag{3.4}$$

若 V_j 与 W_j 不独立，并且 $0 < P(V_1 > V_2) < 1$，则

$$E(W_1 - W_2 | V_1 > V_2) > E(W_1 - W_2), \ E(W_1 - W_2 | V_1 \leqslant V_2) < E(W_1 - W_2)$$

在多元秩-序模型中，令

$$\boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta} = W_1, \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta} = W_2, \varepsilon_h = \delta_1, \varepsilon_j = \delta_2, y_h^* = V_1, y_j^* = V_2, h \neq j, h、j = 1, 2, \cdots, p$$

若有 $y_h > y_j$，由引理 3.1.2 可得

$$E(\boldsymbol{Z}_h^{\mathrm{T}} - \boldsymbol{Z}_j^{\mathrm{T}} | y_h > y_j)\boldsymbol{\beta} > 0$$

再根据引理 3.1.1 可知，存在常数 k_{hj} 使得

$$E(\boldsymbol{Z}_h^{\mathrm{T}} - \boldsymbol{Z}_j^{\mathrm{T}} | y_h > y_j) = E(\boldsymbol{Z}_h^{\mathrm{T}} - \boldsymbol{Z}_j^{\mathrm{T}} | y_h^* - y_j^* > 0) = k_{hj}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{hj}$$

所以有

$$k_{hj}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\Sigma}_{hj}\boldsymbol{\beta} > 0 。$$

又因为 $\boldsymbol{\Sigma}_{hj}$ 为正定阵，得 $k_{hj} > 0$。

综上所述，可得到定理 3.1.1。

定理 3.1.1 在多元秩-序模型中，若

$$0 < p_{hj} = P(y_h > y_j) < 1, h \neq j, h、j = 1, 2, \cdots, p$$

并且 $\boldsymbol{X}_h, \boldsymbol{X}_j$ $(h \neq j, h、j = 1, 2, \cdots, p)$ 满足线性条件 A1，则有

$$\boldsymbol{\Sigma}_{hj}^{-1}E(\boldsymbol{Z}_h - \boldsymbol{Z}_j | y_h > y_j) = k_{hj}\boldsymbol{\beta}, h \neq j, h、j = 1, 2, \cdots, p \tag{3.5}$$

并且 $k_{hj} > 0$。

3.2　回归系数的估计

如果 $\boldsymbol{X}_j\ (j=1,\cdots,p)$ 满足线性条件 A1，利用定理 3.1.1 可以得到一个估计，这个估计与 $\boldsymbol{\beta}$ 只相差一个正常数因子。在多元秩-序模型中共有 $p(p-1)$ 个这样的估计。为了利用数据包含的全部信息，根据它们都与 $\boldsymbol{\beta}$ 只相差一个正常数因子的性质，可以对这 $p(p-1)$ 个估计求加权平均数。用 $\omega_{hj}\ (h\neq j, h、j=1,\cdots,p)$ 表示权，则加权和的总体形式为

$$\sum_{h\neq j}\omega_{hj}\boldsymbol{\Sigma}_{hj}^{-1}E(\boldsymbol{Z}_h-\boldsymbol{Z}_j \mid y_h>y_j) \tag{3.6}$$

式中，$0\leqslant\omega_{hj}\leqslant 1$，$\sum\limits_{h\neq j}$ 表示对所有的 $h、j=1,\cdots,p, h\neq j$ 的项求和。权 ω_{hj} 和概率 $p_{hj}=P(y_h>y_j)$ 是一种自然的选择方式，式（3.6）可化为

$$\boldsymbol{\gamma}=\sum_{h\neq j}\boldsymbol{\Sigma}_{hj}^{-1}E[(\boldsymbol{Z}_h-\boldsymbol{Z}_j)I(y_h>y_j)]=k\boldsymbol{\beta},\ \ k>0 \tag{3.7}$$

式中，$I(y_h>y_j)$ 为示性函数，当 $y_h>y_j$ 时，取值为 1，其他为 0。

若记 $\boldsymbol{\mu}_{hj}=E[(\boldsymbol{Z}_h-\boldsymbol{Z}_j)I(y_h>y_j)]\ (h\neq j, h、j=1,\cdots,p)$，$(\boldsymbol{Y}^{(i)},\boldsymbol{X}^{(i)})(i=1,\cdots,n)$ 为 n 个 i.i.d.样本，$\boldsymbol{Y}^{(i)}=(y_1^{(i)},\cdots,y_p^{(i)})$，则，$\boldsymbol{\mu}_{hj}$ 的矩估计为

$$\hat{\boldsymbol{\mu}}_{hj}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_h^{(i)}-\hat{\boldsymbol{a}}_h-\boldsymbol{X}_j^{(i)}+\hat{\boldsymbol{a}}_j)\ I(y_h^{(i)}>y_j^{(i)}), h\neq j, h、j=1,\cdots,p$$

式中，$\hat{\boldsymbol{a}}_j=\frac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{X}_j^{(i)}(j=1,\cdots,p)$。那么式（3.7）的样本估计为

$$\hat{\boldsymbol{\gamma}}=\sum_{h\neq j}\hat{\boldsymbol{\Sigma}}_{hj}^{-1}\hat{\boldsymbol{\mu}}_{hj} \tag{3.8}$$

其中

$$\hat{\boldsymbol{\Sigma}}_{hj}=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_h^{(i)}-\hat{\boldsymbol{a}}_h-\boldsymbol{X}_j^{(i)}+\hat{\boldsymbol{a}}_j)^{\mathrm{T}}(\boldsymbol{X}_h^{(i)}-\hat{\boldsymbol{a}}_h-\boldsymbol{X}_j^{(i)}+\hat{\boldsymbol{a}}_j)$$

本书称估计式（3.8）为多元秩-序模型的逆回归估计（Inverse Regression Estimation，IRE）。与 Duan 和 Li[34]提出的逆回归估计比较，式（3.8）是一个显式表达式，避免了对一个估计矩阵求特征值和特征向量，并且与 $\boldsymbol{\beta}$ 只相差一个正常数因子。

3.3 相合性

本节将证明 IRE 是 $\sqrt{n}$ 相合的。

引理 3.3.1 若矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{A}+\boldsymbol{B}$ 是均为可逆方阵，则有

$$(\boldsymbol{A}+\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}$$

式中，$\boldsymbol{I}$ 为单位阵。

证明： 直接验证

$$\begin{aligned}
&(\boldsymbol{A}+\boldsymbol{B})\left[\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}\right]\\
&=(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{A}^{-1}-(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}\\
&=\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A}^{-1}-\left[\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}\right]\\
&=\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A}^{-1}-\left[\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}+\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B}-\boldsymbol{I})(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}\right]\\
&=\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A}^{-1}-\left[\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}+\boldsymbol{B}\boldsymbol{A}^{-1}-\boldsymbol{B}(\boldsymbol{I}+\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{A}^{-1}\right]\\
&=\boldsymbol{I}
\end{aligned}$$

可知结论成立。

证毕。

定理 3.3.1 多元秩-序模型若满足定理 3.1.1 的条件，并且 $\boldsymbol{X}_j\ (j=1,\cdots,p)$ 存在有限的四阶矩，则由式（3.8）给出的估计是 $\sqrt{n}$ 相合的，并且是强相合的。

证明： 因为 $\hat{\boldsymbol{\mu}}_{hj}$ 和 $\hat{\boldsymbol{\Sigma}}_{hj}$ 分别是 $\boldsymbol{\mu}_{hj}$ 和 $\boldsymbol{\Sigma}_{hj}$ 的矩估计，所以有

$$\hat{\boldsymbol{\mu}}_{hj}=\boldsymbol{\mu}_{hj}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)$$

$$\hat{\boldsymbol{\Sigma}}_{hj}=\boldsymbol{\Sigma}_{hj}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right),\ h\neq j,\ h、j=1,2,\cdots,p$$

由引理 3.3.1

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{hj}^{-1}&=\left[\boldsymbol{\Sigma}_{hj}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\right]^{-1}\\
&=\boldsymbol{\Sigma}_{hj}^{-1}-\boldsymbol{\Sigma}_{hj}^{-1}\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\left[\boldsymbol{I}+\boldsymbol{\Sigma}_{hj}^{-1}\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\right]^{-1}\boldsymbol{\Sigma}_{hj}^{-1}\\
&=\boldsymbol{\Sigma}_{hj}^{-1}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)
\end{aligned}$$

因此

$$\begin{aligned}\hat{\gamma} &= \sum_{h\neq j}\left[\boldsymbol{\Sigma}_{hj}^{-1}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\right]\left[\boldsymbol{\mu}_{hj}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\right]\\ &= \sum_{h\neq j}\boldsymbol{\Sigma}_{hj}^{-1}\boldsymbol{\mu}_{hj}+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\\ &= \gamma+\boldsymbol{O}\left(n^{-\frac{1}{2}}\right)\end{aligned}$$

证毕。

3.4　模拟研究

本节考虑两个模拟例子，第一个例子中的扰动项服从正态分布，这时现有的模拟极大似然估计方法可以使用，我们将本书所述的方法和模拟极大似然估计进行比较。第二个例子中的扰动项不服从正态分布，本节给出了本书所述方法估计的模拟结果。

3.4.1　模拟研究一

取自变量 $\boldsymbol{X}$ 为 4×4 的随机矩阵，$\mathrm{vec}(\boldsymbol{X})\sim N_{16}(\boldsymbol{\mu},\boldsymbol{\Sigma})$。其中，$\mathrm{vec}(\boldsymbol{X})$ 表示将矩阵 $\boldsymbol{X}$ 的各行向量拉直所得的列向量，即

$$\boldsymbol{\mu}=(2,\ 1,\ 1,\ 0.5,-1,\ 1,\ 1,\ 3,-0.5,-1,\ 0,\ 2,\ 2,\ 1,\ 1,\ 3)^{\mathrm{T}}$$

$$\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_1\otimes\boldsymbol{\Sigma}_2$$

式中，符号 $\otimes$ 表示矩阵的 Kronecker 积；

$$\boldsymbol{\Sigma}_1=\begin{bmatrix}1 & 0.5 & -0.5 & 0.2\\ 0.5 & 2 & 0.3 & -0.2\\ -0.5 & 0.3 & 1 & 0.2\\ 0.2 & -0.2 & 0.2 & 2\end{bmatrix};$$

$$\boldsymbol{\Sigma}_2=\begin{bmatrix}2 & -0.4 & 0.3 & 0.5\\ -0.4 & 1 & -0.2 & 0.3\\ 0.3 & -0.2 & 2 & -0.3\\ 0.5 & 0.3 & -0.3 & 1\end{bmatrix}。$$

扰动项 $\boldsymbol{\varepsilon}\sim N_4(\boldsymbol{0},\boldsymbol{\Omega})$ 有

$$\boldsymbol{\Omega}=\begin{bmatrix} 1 & -0.2 & 0.3 & -0.4 \\ -0.2 & 2 & -0.25 & 0.8 \\ 0.3 & -0.25 & 1 & 0.1 \\ -0.4 & 0.8 & 0.1 & 2 \end{bmatrix}$$

回归系数 $\boldsymbol{\beta}=(1,-2,0,1)^{\mathrm{T}}$，样本容量取 20，50，100，200，500 进行 Monte-Carlo 研究，实验重复 500 次。用 $\hat{\gamma}$ 表示 IRE，其各分量均值和标准差列入表 3.1 中，表 3.3 中列出了 $(\hat{\gamma}_2/\hat{\gamma}_1, \hat{\gamma}_3/\hat{\gamma}_1, \hat{\gamma}_4/\hat{\gamma}_1)$ 的均值及标准差。评价方向估计的另一标准为

$$c=\cos(\hat{\gamma},\boldsymbol{\beta})=\frac{\hat{\gamma}^{\mathrm{T}}\boldsymbol{\beta}}{\sqrt{(\hat{\gamma}^{\mathrm{T}}\hat{\gamma})}\sqrt{(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta})}} \tag{3.9}$$

c 值的均值和标准差列入表 3.4 中。同时求解模拟极大似然估计（SMLE）并与本章方法进行比较。这里利用文献[16]中的 GHK 模拟方法计算概率 $P_{Y^{(i)}}(\boldsymbol{X}^{(i)};\boldsymbol{\beta},\boldsymbol{\Omega})$。GHK 模拟的随机模拟次数 R 与文献[16]中的相同，取 R=500。用 $\hat{\boldsymbol{\beta}}$ 表示 SMLE 中回归系数的估计，其结果列入表 3.2、表 3.3 和表 3.4 中。

表 3.1　IRE 估计的均值及各分量的标准差（括号内为各分量的标准差）

n \ $\hat{\gamma}$		$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$
$n=20$	均值	2.0937	−4.0833	0.1328	2.2528
	标准差	（0.6294）	（0.8795）	（0.4755）	（0.8795）
$n=50$	均值	2.0279	−4.0069	0.0220	1.9975
	标准差	（0.3907）	（0.5757）	（0.2571）	（0.4838）
$n=100$	均值	1.9808	−3.8862	−0.0206	1.9067
	标准差	（0.2995）	（0.4405）	（0.1778）	（0.3699）
$n=200$	均值	1.9117	−3.8709	−0.0099	1.9452
	标准差	（0.1848）	（0.3303）	（0.1244）	（0.2245）
$n=500$	均值	1.9235	−3.8144	0.0056	1.9135
	标准差	（0.1208）	（0.2751）	（ 0.0806）	（0.1321）

表 3.2　SMLE 估计的均值及各分量的标准差（括号内为各分量的标准差）

n \ $\hat{\beta}$		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$n=20$	均值	1.2285	−2.4788	0.0329	1.2468
	标准差	（0.3387）	（0.5812）	（0.2279）	（0.5043）
$n=50$	均值	1.1637	−2.2968	−0.0173	1.2125
	标准差	（0.2371）	（0.4265）	（0.1363）	（0.2480）

续表

n \ $\hat{\beta}$		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$n=100$	均值	1.0179	−2.1734	−0.0124	1.0864
	标准差	（0.1308）	（0.2544）	（0.0913）	（0.1338）
$n=200$	均值	1.0361	−2.0407	−0.0097	1.0839
	标准差	（0.0819）	（0.1754）	（0.0650）	（0.0879）
$n=500$	均值	1.0290	−2.0398	−0.0075	1.0673
	标准差	（0.0527）	（0.1179）	（0.0631）	（0.0528）

表 3.3　方向的均值及各分量的标准差（一）（括号内为各分量的标准差）

样本容量 \ 估计		IRE			SMLE		
		$\hat{\gamma}_2/\hat{\gamma}_1$	$\hat{\gamma}_3/\hat{\gamma}_1$	$\hat{\gamma}_4/\hat{\gamma}_1$	$\hat{\beta}_2/\hat{\beta}_1$	$\hat{\beta}_3/\hat{\beta}_1$	$\hat{\beta}_4/\hat{\beta}_1$
$n=20$	均值	−2.1553	0.0765	1.2218	−2.1215	0.0252	1.0325
	标准差	（1.0824）	（0.4225）	（1.0504）	（0.4039）	（0.1758）	（0.3636）
$n=50$	均值	−2.0576	0.0172	1.0371	−2.0857	−0.0233	1.0835
	标准差	（0.5766）	（0.1807）	（0.3793）	（0.3462）	（0.1017）	（0.3048）
$n=100$	均值	−2.0056	−0.0062	0.9899	−2.0932	−0.0107	1.0645
	标准差	（0.3287）	（0.0823）	（0.2084）	（0.1849）	（0.0840）	（0.1135）
$n=200$	均值	−2.0248	−0.0028	1.0054	−2.0432	−0.0077	1.0326
	标准差	（0.1633）	（0.0612）	（0.1338）	（0.1195）	（0.0637）	（0.0725）
$n=500$	均值	−1.9956	0.0017	0.9986	−2.0187	−0.0035	1.0310
	标准差	（0.1161）	（0.0412）	（0.0765）	（0.0878）	（0.0606）	（0.0449）

表 3.4　c 值及标准差（一）（括号内为标准差）

IRE					
n	20	50	100	200	500
c	0.9637	0.9882	0.9950	0.9981	0.9991
标准差	（0.0434）	（0.0119）	（0.0051）	（0.0020）	（0.0008）
SMLE					
n	20	50	100	200	500
c	0.9821	0.9912	0.9962	0.9985	0.9992
标准差	（0.0225）	（0.0098）	（0.0036）	（0.0015）	（0.0006）

由于 IRE 是一个方向估计，可以相差一个正常数因子，所以分析表 3.1 和表 3.2 直接的模拟结果不方便，可看表 3.3 和表 3.4 中的方向评价指标数据。其中，在 $n\geqslant 50$ 时，表 3.3 的 IRE 均值更接近真值，这说明 IRE 方向的无偏性好于 SMLE 方向。在其他方面，SMLE 要优于 IRE，这是由于 SMLE 是参数方法

而 IRE 是非参数方法，所以 SMLE 的估计效率比较高。但在取大样本时，IRE 的估计效果已很接近 SMLE 了，这说明 IRE 是 $\sqrt{n}$ 相合的，并且有良好的大样本性质。IRE 还有计算量小、节省计算时间等优点。

3.4.2 模拟研究二

多元秩-序模型 SMLE 要求扰动项服从正态分布，而 IRE 只要求扰动项与自变量独立，对扰动项的分布无其他限制，因此 IRE 可以解决扰动项为非正态分布的估计问题。在模拟研究二中，自变量与模拟研究一中的相同，向量 $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)^{\mathrm{T}}$ 与 $\boldsymbol{X}$ 独立并服从球面上的均匀分布，随机数由下面的方式产生：4 维随机向量 ξ 的每一个分量都独立并且服从 $N(0, 1)$ 分布，$(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)^{\mathrm{T}} = 1.5 \times \dfrac{\xi}{\|\xi\|}$，其中 $\|\xi\|$ 为向量 ξ 的模。计算 IRE 的相关指标并填入表 3.5～表 3.7 中。

表 3.5　估计的均值及各分量的标准差（括号内为各分量的标准差）

n	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	$\hat{\gamma}_4$
20	2.0893（0.8505）	−4.2834（1.2204）	−0.0459（0.7141）	2.2044（0.8926）
50	2.0526（0.3765）	−4.1514（0.6598）	−0.0561（0.2937）	2.1307（0.5202）
100	2.0882（0.3060）	−4.0221（0.4722）	−0.0063（0.2210）	2.0019（0.3621）
200	2.0106（0.2277）	−4.0358（0.3031）	0.0101（0.1204）	1.9966（0.2556）
500	2.0272（0.1206）	−4.0301（0.2942）	−0.0205（0.0132）	2.0327（0.1479）

表 3.6　方向的均值及各分量的标准差（二）（括号内为各分量的标准差）

n	$\hat{\gamma}_2/\hat{\gamma}_1$	$\hat{\gamma}_3/\hat{\gamma}_1$	$\hat{\gamma}_4/\hat{\gamma}_1$
20	−2.3488（1.1531）	−0.0026　（0.6320）	1.2564（1.0867）
50	−2.0982（0.5730）	−0.0195（0.1563）	1.0972（0.3668）
100	−1.9864（0.3069）	0.0020（0.1052）	0.9832（0.2236）
200	−2.0243（0.1716）	−0.0063（0.0688）	1.0043（0.1537）
500	−1.9941（0.1090）	−0.0039（0.0382）	1.0066（0.0883）

表 3.7　c 值及标准差（二）（括号内为标准差）

n	20	50	100	200	500
c	0.9625	0.9875	0.9949	0.9980	0.9990
标准差	（0.0485）	（0.0128）	（0.0051）	（0.0021）	（0.0009）

模拟研究二的结果与模拟研究一的结果相似，说明当样本量较大、扰动项为非正态分布时，IRE 也有很好的表现。因此当扰动项为非正态分布时，用 IRE 估计回归系数是合理的。

3.5　回归系数的 Bootstrap 检验

本章给出的回归系数估计与第 2 章的估计有相近的形式，也是一种矩估计函数形式，因此可以断言，该估计也应该是渐近正态的。但是，因为给出的逆回归估计中分片过多，共有 $p(p-1)$个分片估计的和，所以第 2 章中的δ-方法很难得到估计的渐近分布。对分布近似的方法除了利用极限分布之外，还有一类常用的方法，就是 Bootstrap 方法。自从 Efron[65]提出 Bootstrap 方法后，该方法就广泛地应用于统计的各个分支。利用 Bootstrap 方法构造的统计量分布估计可以作为参数的区间估计、假设检验等。

3.5.1　回归系数的线性假设检验

本节考虑应用 Bootstrap 方法得到回归系数的线性假设检验

$$H_0\colon\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=0\ \leftrightarrow\ H_1\colon\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}\neq 0 \tag{3.10}$$

式中，$\boldsymbol{L}$ 为已知的$q\times r$ 的列满秩矩阵。这等价于

$$H_0\colon\ \boldsymbol{L}^{\mathrm{T}}\gamma=0\ \leftrightarrow\ H_1\colon\ \boldsymbol{L}^{\mathrm{T}}\gamma\neq 0 \tag{3.11}$$

式中，γ 为式（3.7）给出的形式。

假设检验就是确定拒绝域，记$\left[(\boldsymbol{X}^{(1)},\ \boldsymbol{Y}^{(1)}),\cdots,(\boldsymbol{X}^{(n)},\ \boldsymbol{Y}^{(n)})\right]$是 n 个来自多元秩-序模型的 i.i.d.样本，并且它们的联合分布为 $F^{(n)}$，令

$$\boldsymbol{R}_n=n(\hat{\gamma}-\gamma)^{\mathrm{T}}\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)$$

式中，$\hat{\gamma}_n$ 为式（3.8）给出的估计，γ 为参数真值。若有

$$\boldsymbol{R}_n=\boldsymbol{R}_n\left[(\boldsymbol{X}^{(1)},\ \boldsymbol{Y}^{(1)}),\cdots,(\boldsymbol{X}^{(n)},\ \boldsymbol{Y}^{(n)}),\ F_n\right]$$

的分布为 G_n，令检验式（3.10）的检验统计量为

$$T_n=n\hat{\gamma}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}\hat{\gamma}$$

则检验式（3.10）显著性水平为α 检验的拒绝域为

$$C=\left\{(\boldsymbol{X},\ \boldsymbol{Y}),T_n>G_n^{-1}(1-\alpha)\right\} \tag{3.12}$$

式中，$G_n^{-1}(1-\alpha)$ 表示分布 G_n 的 $(1-\alpha)$ 下分位数。

在一个给定问题中发现一个具有精确已知分布的检验统计量是很困难的，也就是说，很难找到一个 $\boldsymbol{R}_n$ 具有已知的分布 G_n。如果 G_n 未知，则必须求 G_n 的近似。在传统的近似方法中，可以利用 G_n 的极限分布代替 G_n。而本节不仅精确分布未知，而且也无法得到其极限分布，所以可以考虑应用 Bootstrap 方法。

这里利用 $\boldsymbol{R}_n^* = n(\gamma^* - \hat{\gamma})^{\mathrm{T}} \boldsymbol{LL}^{\mathrm{T}} (\gamma^* - \hat{\gamma})$ 的分布 G_B 来渐近 $\boldsymbol{R}_n = n(\hat{\gamma} - \gamma)^{\mathrm{T}} \boldsymbol{LL}^{\mathrm{T}} (\hat{\gamma} - \gamma)$ 的分布，其中 $\hat{\gamma}^*$ 为 γ 的 Bootstrap 估计。其步骤如下：

（1）在原样本中随机抽取 n_1 个样本，计算每次回归系数的 Bootstrap 估计 γ^* 与 $\boldsymbol{L}^{\mathrm{T}}\gamma^*$。

（2）计算 $\boldsymbol{R}_n^*$，并重复 B 次，对于给定的显著性水平 α，令 $c_{(1-\alpha)}^*$ 为 $\boldsymbol{R}_n^*$ 的第 $[B \times (1-\alpha) + 1]$ 个顺序统计量，则显著性水平为 α 的 Bootstrap 检验的拒绝域为

$$C^* = \left\{ (\boldsymbol{X}, \boldsymbol{Y}), T_n > c_{(1-\alpha)}^* \right\} \tag{3.13}$$

下面给出检验的渐近理论。

首先，给出一些关于 Bootstrap 检验的定义。令 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ 为 n 个来自随机模型 P_n 的 p 维向量观测，并且令

$$\boldsymbol{R}_n = \boldsymbol{R}_n(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n, P_n)$$

为一个 s 维随机向量。记

$$H_n(x) = P\{\boldsymbol{R}_n \leqslant \boldsymbol{x} \mid \boldsymbol{P}_n\}$$

$H_n(x)$ 的 Bootstrap 估计为

$$H_B(x) = P_*\{\boldsymbol{R}_n^* \leqslant \boldsymbol{x} \mid \hat{\boldsymbol{P}}_n\}$$

式中，$\boldsymbol{R}_n^* = \boldsymbol{R}_n(\boldsymbol{X}_1^*, \cdots, \boldsymbol{X}_n^*, \hat{P}_n)$， $\boldsymbol{X}_1^*, \cdots, \boldsymbol{X}_n^*$ 是由估计模型 $\hat{P}_n$ 中抽取的一个 Bootstrap 样本， $P_*\{\cdot \mid \hat{P}_n\}$ 为给定 $\hat{P}_n$ 的条件概率。

令 $\mathcal{F}^{(n)}$ 为 $F^{(n)}$ 的所有可能分布的集合。$\mathcal{F}_0^{(n)}$ 和 $\mathcal{F}_1^{(n)}$ 为 $\mathcal{F}^{(n)}$ 的两个不相交子集，检验

$$\boldsymbol{H}_0: F^{(n)} \in \mathcal{F}_0^{(n)} \quad \leftrightarrow \quad H_1: F^{(n)} \in \mathcal{F}_1^{(n)}$$

定义 3.4.1 令 α 是一个给定的名义水平，拒绝域为 C_n 的检验如果有

$$\lim_{n \to \infty} P\left\{ (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n) \in C_n \mid F^{(n)} \right\} = \alpha$$

对于任意 $F^{(n)} \in \mathcal{F}_0^{(n)}$ 都成立，则称检验是渐近精确的（Asymptotically Correct）。

定义 3.4.2 令 α 是一个给定的名义水平，如果拒绝域为 C_n 的检验有

$$\lim_{n \to \infty} P\left\{ (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n) \in C_n \mid F^{(n)} \right\} = 1$$

对于任意 $F^{(n)} \in \mathcal{F}_1^{(n)}$ 成立，则称检验是相合的。

定义 3.4.3 令 $\mathcal{F}_{\boldsymbol{R}^p}$ 表示 $\boldsymbol{R}^p$ 上的所有分布的集合， ρ 为定义在 $\mathcal{F}_{\boldsymbol{R}^p}$ 上的一个度量，如果当 $n \to \infty$ 时有 $\rho(H_B, H_{P_n}) \to_p 0$， 则称 H_B 是 ρ 相合的；如果当 $n \to \infty$ 时有 $\rho(H_B, H_{P_n}) \to_{a \cdot s \cdot 0}$， 则称 H_B 是 ρ 强相合的。

引理 3.4.1 令 $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$ 为 i.i.d.随机向量，假设 $E \| \boldsymbol{X}_1 \|^2 < \infty$，$T_n = g(\bar{\boldsymbol{X}}_n)$，

$\boldsymbol{\theta}=\boldsymbol{g}(\boldsymbol{\mu})$，其中 $\bar{\boldsymbol{X}}_n=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i$， $\boldsymbol{\mu}=E\boldsymbol{X}_1$。令 $\boldsymbol{R}_n=\sqrt{n}(\boldsymbol{T}_n-\boldsymbol{\theta})$，若 $\boldsymbol{g}(\cdot)$ 在 $\boldsymbol{\mu}=E\boldsymbol{X}_1$ 处连续可导，并且有 $\nabla\boldsymbol{g}(\boldsymbol{\mu})\neq 0$， $\nabla\boldsymbol{g}(\boldsymbol{\mu})$ 表示函数 $\boldsymbol{g}$ 在 $\boldsymbol{\mu}$ 处的梯度，则 H_n 的 Bootstrap 估计 H_B 是强相合的，即 $\rho(H_B,H_n)\to_{a.s.}0$。

引理 3.4.1 是文献[65]中的定理 3.1。

在多元秩-序模型中，若记

$$\begin{aligned}\boldsymbol{U}^{(i)}=\Big[&(\boldsymbol{X}_1^{(i)}-\boldsymbol{X}_2^{(i)})^{\mathrm{T}}\boldsymbol{I}(y_1^{(i)}>y_2^{(i)}),\cdots,(\boldsymbol{X}_1^{(i)}-\boldsymbol{X}_p^{(i)})^{\mathrm{T}}\boldsymbol{I}(y_1^{(i)}>y_p^{(i)}),\cdots,\\&(\boldsymbol{X}_p^{(i)}-\boldsymbol{X}_1^{(i)})^{\mathrm{T}}\boldsymbol{I}(y_p^{(i)}>y_1^{(i)}),\cdots,(\boldsymbol{X}_p^{(i)}-\boldsymbol{X}_{p-1}^{(i)})^{\mathrm{T}}\boldsymbol{I}(y_p^{(i)}>y_{p-1}^{(i)}),\\&\boldsymbol{X}_1^{(i)\mathrm{T}},\cdots,\boldsymbol{X}_p^{(i)\mathrm{T}},\ \mathrm{vec}(\boldsymbol{X}_1^{(i)}\boldsymbol{X}_1^{(i)\mathrm{T}})^{\mathrm{T}},\cdots,\ \mathrm{vec}(\boldsymbol{X}_p^{(i)}\boldsymbol{X}_p^{(i)\mathrm{T}})^{\mathrm{T}}\Big]^{\mathrm{T}}\end{aligned}$$

则 $\boldsymbol{U}^{(i)}$ 为 i.i.d.随机向量，再令

$$\bar{\boldsymbol{U}}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{U}^{(i)}$$

由第 2 章中估计的渐近正态性证明过程可知，式（3.8）给出的估计是 $\bar{\boldsymbol{U}}$ 的函数，这个函数在 $E(\boldsymbol{U}^{(1)})$ 处是连续可导的，并且导数不为 0。因此，可以得到定理 3.4.1。

定理 3.4.1　在多元秩-序模型中，若 $\boldsymbol{X}_j$ 存在有限的四阶矩，则 $\sqrt{n}(\gamma^*-\hat{\gamma})$ 分布的 Bootstrap 估计 H_B 关于 $\sqrt{n}(\hat{\gamma}-\gamma)$ 的分布 H_n 是强相合的。进一步有

$$\boldsymbol{R}_n^*=n(\gamma^*-\hat{\gamma})^{\mathrm{T}}\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}(\gamma^*-\hat{\gamma})$$

分布的 Bootstrap 估计 G_B 关于 $\boldsymbol{R}_n=n(\hat{\gamma}-\gamma)^{\mathrm{T}}\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)$ 的分布 G_n 是强相合的。

由定理 3.4.1 中给出分布 Bootstrap 估计的相合性，可以得到定理 3.4.2。

定理 3.4.2　在多元秩-序模型中，若 $\boldsymbol{X}_j$ 存在有限的四阶矩，对于给定的水平 $\alpha,0<\alpha<1$，则由式（3.13）给出拒绝域的检验，是式（3.11）的渐近精确的检验。

证明：当假设检验式（3.11）的原假设成立时，有

$$\begin{aligned}&P\Big\{\Big[(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}),\cdots,(\boldsymbol{X}^{(n)},\boldsymbol{Y}^{(n)})\Big]\in C^*\mid F^{(n)}\Big\}\\&=P[T_n>G_B^{-1}(1-\alpha)\mid F^{(n)}]\\&=P[n\hat{\gamma}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}\hat{\gamma}>G_B^{-1}(1-\alpha)\mid F^{(n)}]\\&=P[n(\hat{\gamma}^{\mathrm{T}}-\gamma^{\mathrm{T}})\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)>G_B^{-1}(1-\alpha)\mid F^{(n)}]\\&=P[n(\hat{\gamma}^{\mathrm{T}}-\gamma^{\mathrm{T}})\boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)>G_n^{-1}(1-\alpha)]+o(1)\\&=1-G_n[G_n^{-1}(1-\alpha)]+o(1)\to\alpha\end{aligned}$$

证毕。

定理 3.4.3 在多元秩-序模型中，$\boldsymbol{X}_j$ 存在有限的四阶矩，若有 $n^{1/2}\|\boldsymbol{L}^{\mathrm{T}}\gamma\|\to\infty$，则由式（3.13）给出拒绝域的检验是式（3.11）的相合检验。

证明： 因为 $\hat{\gamma}\to_{a.s.}\gamma$，若有 $n^{1/2}\|\boldsymbol{L}^{\mathrm{T}}\gamma\|\to\infty$，则对应的 $M>0$，有 $\lim\limits_{n\to\infty}P(T_n>M)=1$。而 $\hat{\gamma}$ 是 $\sqrt{n}$ 相合的，这样，当 $\boldsymbol{X}_j$ 存在有限的四阶矩时，$\boldsymbol{R}_n=n(\hat{\gamma}^{\mathrm{T}}-\gamma^{\mathrm{T}})\ \ \boldsymbol{L}\boldsymbol{L}^{\mathrm{T}}(\hat{\gamma}-\gamma)$ 的分布 G_n 的 $1-\alpha$ 分位数 $G_n^{-1}(1-\alpha)<\infty$。所以有

$$\begin{aligned}&P\left\{\left[(\boldsymbol{X}^{(1)},\boldsymbol{Y}^{(1)}),\cdots,(\boldsymbol{X}^{(n)},\boldsymbol{Y}^{(n)})\right]\in C^*\mid F^{(n)}\right\}\\&=P[T_n>G_B^{-1}(1-\alpha)\mid F^{(n)}]\\&=P[T_n>G_n^{-1}(1-\alpha)]+o(1)\to1\end{aligned}$$

证毕。

3.5.2 假设检验的模拟实验

本节通过模拟研究当样本容量有限时，3.4.1 节给出的 Bootstrap 检验的水平和势。

模拟中仍然取 3.4 节中模拟研究二的模型产生随机数，首先模拟犯第一类错误的概率，矩阵 $\boldsymbol{L}$ 取为

$$\boldsymbol{L}^{\mathrm{T}}=\begin{bmatrix}1&1&1&1\\0&1&0&2\end{bmatrix}$$

在模拟过程中，样本容量取 $n=100,\ 300,\ 500$。在用 Bootstrap 估计 $G_B^{-1}(1-\alpha)$ 时，每个 Bootstrap 数据集抽取的样本量 $n_1=n$，Bootstrap 重复次数 $B=5000$，检验重复 5000 次，得到犯第一类错误的频率，将模拟结果列入表 3.8 中。

表 3.8 检验犯第一类错误的频率

n	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.025$
100	0.078	0.038	0.020
300	0.096	0.046	0.024
500	0.098	0.052	0.026

从模拟结果来看，当样本容量为 $n=100$ 时，犯第一类错误的频率会小于名义水平，而随着样本容量的增大，当 $n=300$ 和 $n=300$ 时会趋近于真实水平，说明本节给出的 Bootstrap 检验是渐近精确的。但是 Bootstrap 检验在样本容量较小时，检验水平会使用不足。

下面我们利用计算机模拟来考察上述检验的势。回归系数 $\boldsymbol{\beta}$ 的真值分别取以下不同的值。$\boldsymbol{\beta}=(1,-2.25,\ 0,\ 1)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1,-2,\ 0,1.25)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1,-2.5,\ 0,\ 1)^{\mathrm{T}}$,

$\boldsymbol{\beta}=(1,-2,\ 0,\ 1.5)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1,-1,\ 0,\ 1)^{\mathrm{T}}$， $\boldsymbol{L}$ 的取值仍为

$$\boldsymbol{L}^{\mathrm{T}}=\begin{bmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 0 & 2\end{bmatrix}$$

满足

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=\begin{bmatrix}-0.25\\ -0.25\end{bmatrix},\begin{bmatrix}0.25\\ 0.5\end{bmatrix},\begin{bmatrix}-0.5\\ -0.5\end{bmatrix},\begin{bmatrix}0.5\\ 1\end{bmatrix},\begin{bmatrix}1\\ 1\end{bmatrix}$$

模拟结果如表 3.9 所示。

表 3.9　检验的势（一）（ α =0.05）

$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	$(-0.25，-0.25)^{\mathrm{T}}$	$(0.25，0.5)^{\mathrm{T}}$	$(-0.5，-0.5)^{\mathrm{T}}$	$(0.5，1)^{\mathrm{T}}$	$(1，1)^{\mathrm{T}}$
n=100	0.1922	0.2680	0.3288	0.5736	0.9734
n=300	0.2954	0.4318	0.5964	0.9602	1
n=500	0.4636	0.7159	0.8710	0.9982	1
n=1000	0.6742	0.9840	0.9926	1	1

下面考察 $\boldsymbol{L}$ 取值不同时，不同检验的势。令 $\boldsymbol{L}^{\mathrm{T}}$ 为下列矩阵：

$$\begin{bmatrix}1 & 1 & 1 & 1.5\\ 0 & 1 & 0 & 2\end{bmatrix},\begin{bmatrix}1 & 1 & 1 & 1\\ 0 & 1.25 & 0 & 2\end{bmatrix},\begin{bmatrix}1 & 1 & 1 & 1.5\\ 0 & 1.25 & 0 & 2\end{bmatrix}$$

$$\begin{bmatrix}1 & 1 & 1 & 1\\ 0 & 1.5 & 0 & 2\end{bmatrix},\begin{bmatrix}1 & 1 & 1 & 2\\ 0 & 1.5 & 0 & 2\end{bmatrix}$$

使得

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=\begin{bmatrix}0.5\\ 0\end{bmatrix},\begin{bmatrix}0\\ -0.5\end{bmatrix},\begin{bmatrix}0.5\\ -0.5\end{bmatrix},\begin{bmatrix}0\\ -1\end{bmatrix},\begin{bmatrix}1\\ -1\end{bmatrix}$$

令检验的显著性水平为 0.05，计算检验的势，结果列入表 3.10 中。

表 3.10　检验的势（二）（ $\alpha=0.05$）

$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	$(0.5，0)^{\mathrm{T}}$	$(0，-0.5)^{\mathrm{T}}$	$(0.5，-0.5)^{\mathrm{T}}$	$(0，-1)^{\mathrm{T}}$	$(1，-1)^{\mathrm{T}}$
n=100	0.2884	0.2172	0.5418	0.5082	0.8974
n=300	0.6626	0.5684	0.8672	0.8490	0.9996
n=500	0.8584	0.7324	0.9842	0.9694	1
n=1000	0.9956	0.9918	0.9996	0.9992	1

第 4 章　广义多项选择模型回归系数的估计

4.1　广义多项选择模型

SIR 方法的目的是从高维解释变量中寻找低维的解释变量，其优点是不依赖于响应变量与解释变量之间的函数关系，不依赖于扰动项的分布，并且计算简单。MNP 模型是一种特殊的降维模型，与一般的降维模型略有不同，这在下面的介绍中可以看到。若仅对回归系数 $\boldsymbol{\beta}$ 感兴趣，则有两点促使我们采用 SIR 方法来估计 $\boldsymbol{\beta}$：一是前面已指出，实际上 $\boldsymbol{\beta}$ 是不可估的，只能得到 $\boldsymbol{\beta}$ 的方向估计；二是 MNP 模型中映射 $\tau(\cdot)$ 已提供了一个自然的分片方法。

鉴于 SIR 方法的优点，本章将定义 1.1.2 中的多项选择模型推广，假定

$$\boldsymbol{Y}^* = \boldsymbol{g}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}, \boldsymbol{\varepsilon}) \tag{4.1}$$

其中，

$$\boldsymbol{g}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}, \boldsymbol{\varepsilon}) = [f(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta}, \varepsilon_1), f(\boldsymbol{X}_2^{\mathrm{T}}\boldsymbol{\beta}, \varepsilon_2), \cdots, f(\boldsymbol{X}_p^{\mathrm{T}}\boldsymbol{\beta}, \varepsilon_p)]^{\mathrm{T}}$$

函数 f 未知；$\boldsymbol{X}_h^{\mathrm{T}}$ 为 $\boldsymbol{X}^{\mathrm{T}}$ 的第 h 行；ε_h 为 $\boldsymbol{\varepsilon}$ 的第 h 个分量，$h = 1, 2, \cdots, p$；$\boldsymbol{X}$ 与 $\boldsymbol{\varepsilon}$ 独立，$\boldsymbol{\varepsilon}$ 的分布未知，但对 $\boldsymbol{X}$ 有一个线性条件的假定，这个假定当 $\boldsymbol{X}$ 是正态随机矩阵时满足，Y 与 $\boldsymbol{Y}^*$ 的关系仍由式（1.3）给出，即 $Y = \arg\max\limits_h \{y_1^*, y_2^*, \cdots, y_p^*\}$。我们把由式（1.3）和式（4.1）确定的模型称为广义多项选择模型（Generalized Multinomial Choice Model），简称 GMNC 模型。这里无须对 f 的具体形式作任何假定，但根据 Li[35]中的 Remark 4.5，要排除一些对称函数 f，以保证逆回归函数是非平凡的。由于在这里 g 或 f 的形式未知，所以广义多项选择模型的回归系数可以识别到相差一个常数因子，这个常数是可正可负的。

4.2 节研究 $\boldsymbol{X}$ 对 $\boldsymbol{Y}$ 的逆回归性质，在线性条件下，给出条件期望和回归系数 $\boldsymbol{\beta}$ 的关系，根据这个关系给出一个 $\boldsymbol{\beta}$ 的方向估计。4.3 节研究估计的渐近正态性；并在 4.4 节给出模拟结果。

最后，需要指出，本章的方法可适用于其他 LDV 模型。

4.2　广义多项选择模型中回归系数的估计

考虑 4.1 节中的模型式（4.1），记 $\mathrm{vec}(\boldsymbol{X})$ 为矩阵 $\boldsymbol{X}$ 按列拉直得到的 $p\times q$ 维列向量，$\boldsymbol{\mu}=E[\mathrm{vec}(\boldsymbol{X})], \boldsymbol{\Sigma}=\mathrm{Cov}[\mathrm{vec}(\boldsymbol{X})]$，并且要求 $\boldsymbol{\Sigma}>0$，令 $\boldsymbol{Z}=\mathrm{vec}(\boldsymbol{X})-\boldsymbol{\mu}$, $\boldsymbol{B}=\boldsymbol{I}_p\otimes\boldsymbol{\beta}$，这里记号 $\otimes$ 表示矩阵的 Kronecker 积（下同）。因为 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}=\boldsymbol{B}^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X})$，模型可表示为

$$\boldsymbol{Y}^*=g[\boldsymbol{B}^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X}),\ \boldsymbol{\varepsilon}]$$

写成降维模型形式的 GMNC 模型为

$$Y=\tau\{\boldsymbol{g}[\boldsymbol{B}^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X}),\ \boldsymbol{\varepsilon}]\}\tag{4.2}$$

式中，$\tau(\cdot)$ 由式（1.3）确定。模型式（4.2）与 Li[35]中降维模型的主要不同之处是矩阵 $\boldsymbol{B}$ 具有特殊结构。尽管如此，在线性条件下，关于逆回归的基本定理成立。

线性条件：对任意 $\boldsymbol{b}\in\mathbf{R}^{pq}$，条件期望 $E[\boldsymbol{b}^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X})|\ \boldsymbol{B}^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X})]$ 是 $\boldsymbol{B}^{\mathrm{T}}\mathrm{vec}(\boldsymbol{X})$ 的线性函数。

引理 4.2.1 就是逆回归的基本定理。

引理 4.2.1　在 GMNC 模型式（4.2）下，若满足线性条件，则

$$E(\boldsymbol{Z}|\ Y)\in\boldsymbol{S}(\boldsymbol{\Sigma}\boldsymbol{B})$$

式中，$\boldsymbol{S}(\boldsymbol{\Sigma}\boldsymbol{B})$ 表示由矩阵 $\boldsymbol{\Sigma}\boldsymbol{B}$ 的列向量张成的线性空间。

记 $\boldsymbol{e}_j$ 为第 j 个分量为 1，其余为 0 的 p 维列向量，$\boldsymbol{Q}_j=\boldsymbol{e}_j^{\mathrm{T}}\otimes\boldsymbol{I}_q\,(j=1,2,\cdots,p)$。由引理 4.2.1 得，$\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y)\in\boldsymbol{S}(\boldsymbol{B})$，由于 $\boldsymbol{B}=\boldsymbol{I}_p\otimes\boldsymbol{\beta}$，因而存在 k_j（可与 Y 及其参数有关），使得

$$\boldsymbol{Q}_j\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y)=k_j\boldsymbol{\beta}, j=1,2,\cdots,p\tag{4.3}$$

设 $(\boldsymbol{X}^{(i)}, Y^{(i)})(i=1, 2,\cdots, n)$ 为来自 GMNP 模型式（4.2）的 i.i.d.样本，根据关系式（4.3），可给出 $\boldsymbol{\beta}$ 的方向估计。记

$$\boldsymbol{p}_h=P(Y=h)$$

$$\boldsymbol{m}_h=E(\mathrm{vec}(\boldsymbol{X})|\ Y=h)=E(\boldsymbol{Z}|\ Y=h)+\boldsymbol{\mu}, h=1, 2,\cdots, p$$

并令

$$\delta_h^{(i)}=\begin{cases}1,\ Y^{(i)}=h\\0,\ Y^{(i)}\neq h\quad h=1,\cdots,p,\ i=1,\cdots,n\end{cases}$$

则 $\boldsymbol{p}_h$ 和 $\boldsymbol{m}_h$ 的矩估计分别为

$$\hat{\boldsymbol{p}}_h=\frac{1}{n}\sum_{i=1}^{n}\delta_h^{(i)}$$

$$\hat{\boldsymbol{m}}_h = \frac{1}{n\hat{\boldsymbol{p}}_h}\sum_{i=1}^{n}\mathrm{vec}(\boldsymbol{X}^{(i)})\delta_h^{(i)}$$

若取

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{vec}(\boldsymbol{X}^{(i)})$$

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\sum_{i=1}^{n}[\mathrm{vec}(\boldsymbol{X}^{(i)})-\hat{\boldsymbol{\mu}}][\mathrm{vec}(\boldsymbol{X}^{(i)})-\hat{\boldsymbol{\mu}}]^{\mathrm{T}}$$

$\hat{\boldsymbol{\mu}}$和$\hat{\boldsymbol{\Sigma}}$分别为$\boldsymbol{\mu}$和$\boldsymbol{\Sigma}$的矩估计，则$\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y=h)$的一个合理估计为

$$\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}}),\ h=1,\ 2,\cdots,\ p$$

由式（4.3），可得到p^2个$\boldsymbol{\beta}$的方向估计

$$\hat{\boldsymbol{\beta}}_{jh} = \boldsymbol{Q}_j\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}}),\ j、h=1,\ 2,\cdots,\ p \tag{4.4}$$

想要将式（4.4）中的p^2个估计量综合成一个估计，可以构造随机矩阵

$$\hat{\boldsymbol{V}} = \sum_{h=1}^{p}\hat{\boldsymbol{p}}_h\sum_{j=1}^{p}\hat{\boldsymbol{\beta}}_{jh}\hat{\boldsymbol{\beta}}_{jh}^{\mathrm{T}} \tag{4.5}$$

以$\hat{\boldsymbol{V}}$的对应最大特征根的特征向量作为$\boldsymbol{\beta}$的方向估计。现在分析由式(4.5）给出的矩阵$\hat{\boldsymbol{V}}$。由式（4.3）和式（4.4）得

$$(\hat{\beta}_{1h},\hat{\beta}_{2h},\cdots,\hat{\beta}_{ph}) = [\boldsymbol{Q}_1\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}}),\boldsymbol{Q}_2\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}}),\cdots,\boldsymbol{Q}_p\hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}})]$$

其可作为

$$[\boldsymbol{Q}_1\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y=h),\boldsymbol{Q}_2\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y=h),\cdots,\boldsymbol{Q}_p\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y=h)]$$

的估计。所以，$\hat{\boldsymbol{V}}$可作为

$$\begin{aligned}\boldsymbol{V}&=\sum_{h=1}^{p}\boldsymbol{p}_h\sum_{j=1}^{p}\boldsymbol{Q}_j\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y=h)E(\boldsymbol{Z}|\ Y=h)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\\&=E\sum_{j=1}^{p}\boldsymbol{Q}_j\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y)E(\boldsymbol{Z}|\ Y)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\\&=E\boldsymbol{M}\boldsymbol{M}^{\mathrm{T}}\end{aligned} \tag{4.6}$$

的估计，其中$\boldsymbol{M}=[\boldsymbol{Q}_1\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y),\boldsymbol{Q}_2\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y),\cdots,\boldsymbol{Q}_p\boldsymbol{\Sigma}^{-1}E(\boldsymbol{Z}|\ Y)]$。

4.3　渐近正态性

4.2 节得到了$\boldsymbol{\beta}$的方向估计。为了讨论估计的大样本性质，有必要对$\boldsymbol{\beta}$和$\hat{\boldsymbol{\beta}}$做一些假定和限制。令$\boldsymbol{\beta}$为矩阵$\boldsymbol{V}$的非 0 特征值对应的特征向量，$\hat{\boldsymbol{\beta}}$为估计矩阵$\hat{\boldsymbol{V}}$的最大特征值对应的特征向量。因为特征向量可相差一个非 0 常数，所以

对 $\boldsymbol{\beta}$ 做以下假定：

若 $\beta_1 \neq 0$（如果 $\beta_1 = 0$，可以对 β 的其他非 0 分量做下面的假定），

$$\langle \boldsymbol{\beta}, \boldsymbol{\beta} \rangle = 1,\ \beta_1 > 0 \tag{4.7}$$

同时对 $\hat{\boldsymbol{\beta}}$ 做下面的标准化限制：

$$\langle \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}} \rangle = 1,\ \hat{\beta}_1 \geqslant 0 \tag{4.8}$$

首先得到 $\hat{\boldsymbol{V}}$ 的渐近分布。

定理 4.3.1　在 n 个样本的 GMNC 模型中，$\left\{(Y^{(i)}, \boldsymbol{X}^{(i)}),\ i = 1, 2, \cdots, n\right\}$ 为 i.i.d. 样本，并且 $p_h \neq 0(h = 1, \cdots, p)$，在引理 4.2.1 的条件下，当 $\boldsymbol{\Sigma}$ 已知时，若 $E(\boldsymbol{Z}|\ Y) \neq 0$，则有

$$\sqrt{n}(\hat{\boldsymbol{V}} - \boldsymbol{V}) \to_d \boldsymbol{U}_{\boldsymbol{V}}$$

式中，$\to_d$ 表示依分布收敛；$\mathrm{vec}(\boldsymbol{U}_{\boldsymbol{V}}) \sim N(0, \boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}})$，协方差阵 $\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}}$ 为

$$\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}} = \mathrm{Cov}\left(\sum_{j=1}^{p}\left[\boldsymbol{Q}_j \boldsymbol{\Sigma}^{-1} E\left(\boldsymbol{Z}\middle|Y\right)\right]^{[2]}\right) + E\left[\boldsymbol{K}(\boldsymbol{Y})^{\mathrm{T}} \mathrm{Cov}(\boldsymbol{Z}|Y) \boldsymbol{K}(\boldsymbol{Y})\right] \tag{4.9}$$

其中

$$\boldsymbol{K}(\boldsymbol{Y}) = \sum_{j=1}^{p}\left\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}) \otimes [E(\boldsymbol{Z}|\ Y)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}] + [E(\boldsymbol{Z}|\ Y)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}] \otimes (\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\right\}$$

符号 $\boldsymbol{A}^{[2]}$ 表示矩阵 $\boldsymbol{A}$ 的 Kronecker 积 $\boldsymbol{A} \otimes \boldsymbol{A}$。

证明：考虑 $\boldsymbol{\Sigma}$ 已知时的情况，这时式（4.5）给出的 $\boldsymbol{V}$ 的估计为

$$\hat{\boldsymbol{V}} = \sum_{h=1}^{p} \hat{\boldsymbol{p}}_h [\sum_{j=1}^{p} \boldsymbol{Q}_j \boldsymbol{\Sigma}^{-1} (\hat{\boldsymbol{m}}_h - \hat{\boldsymbol{\mu}})(\hat{\boldsymbol{m}}_h - \hat{\boldsymbol{\mu}})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{Q}_j^{\mathrm{T}}]$$

第一步，得到有关矩向量的渐近分布。令 $\boldsymbol{u}_0^{(i)}$ 为 $p + p^2 q + pq$ 维列向量

$$\boldsymbol{u}_0^{(i)} = [\delta_1^{(i)}, \cdots, \delta_p^{(i)}, \mathrm{vec}(\boldsymbol{X}^{(i)})^{\mathrm{T}} \delta_1^{(i)}, \cdots, \mathrm{vec}(\boldsymbol{X}^{(i)})^{\mathrm{T}} \delta_p^{(i)}, \mathrm{vec}(\boldsymbol{X}^{(i)})^{\mathrm{T}}]^{\mathrm{T}}$$

其中，δ_h 为示性函数

$$\delta_h = \begin{cases} 1, & Y = h \\ 0, & Y \neq h \end{cases}$$

由定理的假定知，$\boldsymbol{u}_0^{(i)}$ 为 i.i.d.，存在期望 $\boldsymbol{\mu}_{u_0}$ 和方差 $\boldsymbol{\Sigma}_{u_0}$

$$\boldsymbol{\mu}_{u_0} = (p_1, \cdots, p_p, p_1 \boldsymbol{m}_1^{\mathrm{T}}, \cdots, p_p \boldsymbol{m}_p^{\mathrm{T}}, \boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}}$$

$$\boldsymbol{\Sigma}_{u_0} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{B}_{13} \\ \boldsymbol{B}_{12}^{\mathrm{T}} & \boldsymbol{B}_{22} & \boldsymbol{B}_{23} \\ \boldsymbol{B}_{13}^{\mathrm{T}} & \boldsymbol{B}_{23}^{\mathrm{T}} & \boldsymbol{B}_{33} \end{bmatrix}$$

其中

$$\boldsymbol{B}_{11}=\begin{bmatrix} p_1(1-p_1) & -p_2p_1 & \cdots & -p_pp_1 \\ -p_1p_2 & p_2(1-p_2) & \cdots & -p_pp_2 \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_p & \cdots & \cdots & p_p(1-p_p) \end{bmatrix}$$

$$\boldsymbol{B}_{12}=\begin{bmatrix} p_1(1-p_1)\boldsymbol{m}_1^{\mathrm{T}} & -p_1p_2\boldsymbol{m}_2^{\mathrm{T}} & \cdots & -p_1p_2\boldsymbol{m}_p^{\mathrm{T}} \\ -p_1p_2\boldsymbol{m}_1^{\mathrm{T}} & p_2(1-p_2)\boldsymbol{m}_2^{\mathrm{T}} & \cdots & -p_2p_p\boldsymbol{m}_p^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_1p_p\boldsymbol{m}_1^{\mathrm{T}} & \cdots & \cdots & p_p(1-p_p)\boldsymbol{m}_p^{\mathrm{T}} \end{bmatrix}$$

$$\boldsymbol{B}_{13}=\begin{bmatrix} p_1\boldsymbol{m}_1^{\mathrm{T}}-p_1\boldsymbol{\mu}^{\mathrm{T}} \\ \vdots \\ p_p\boldsymbol{m}_p^{\mathrm{T}}-p_p\boldsymbol{\mu}^{\mathrm{T}} \end{bmatrix}$$

$$\boldsymbol{B}_{22}=\begin{bmatrix} \tilde{V}_1 & -p_1p_2\boldsymbol{m}_1\boldsymbol{m}_2^{\mathrm{T}} & \cdots & -p_1p_p\boldsymbol{m}_1\boldsymbol{m}_p^{\mathrm{T}} \\ -p_2p_1\boldsymbol{m}_2\boldsymbol{m}_1^{\mathrm{T}} & \tilde{\boldsymbol{V}}_2 & \cdots & -p_2p_p\boldsymbol{m}_2\boldsymbol{m}_p^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ -p_pp_1\boldsymbol{m}_p\boldsymbol{m}_1^{\mathrm{T}} & -p_pp_2\boldsymbol{m}_p\boldsymbol{m}_2^{\mathrm{T}} & \cdots & \tilde{\boldsymbol{V}}_p \end{bmatrix}$$

$$\boldsymbol{B}_{23}=\begin{bmatrix} \tilde{\boldsymbol{V}}_1+p_1\boldsymbol{m}_1(p_1\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}} \\ \vdots \\ \tilde{\boldsymbol{V}}_p+p_p\boldsymbol{m}_p(p_p\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}} \end{bmatrix}$$

$$\boldsymbol{B}_{33}=\boldsymbol{\Sigma}$$

$$\tilde{\boldsymbol{V}}_h=E\left[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\delta_h\right]-p_h^2\boldsymbol{m}_h\boldsymbol{m}_h^{\mathrm{T}}$$

令 $\overline{\boldsymbol{u}}_0=\dfrac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{u}_0^{(i)}$，由中心极限定理得

$$\sqrt{n}(\overline{\boldsymbol{u}}_0-\boldsymbol{\mu}_{u_0})\to_d N(\mathbf{0},\boldsymbol{\Sigma}_{u_0})$$

第二步，得到构成 $\hat{\boldsymbol{V}}$ 的随机向量渐近分布，令

$$\boldsymbol{u}_1=(\hat{p}_1,\cdots,\hat{p}_p,\hat{\boldsymbol{m}}_1^{\mathrm{T}},\cdots,\hat{\boldsymbol{m}}_p^{\mathrm{T}},\hat{\boldsymbol{\mu}}^{\mathrm{T}})^{\mathrm{T}}$$

其中，$\hat{p}_h=\sum\limits_{i=1}^{n}\delta_h^{(i)}$, $\hat{\boldsymbol{m}}_h=\dfrac{1}{n}\sum\limits_{i=1}^{n}\mathrm{vec}(\boldsymbol{X}^{(i)})\delta_h^{(i)}/\hat{p}_h$，定义函数

$$\boldsymbol{f}_1:\ \mathbf{R}^{p+p^2q+pq}\quad\to\quad\mathbf{R}^{p+p^2q+pq}$$

$$\boldsymbol{u}=(\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\boldsymbol{c}^{\mathrm{T}})^{\mathrm{T}}\quad\to\quad(\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}}/\boldsymbol{a}_1,\cdots,\boldsymbol{b}_p^{\mathrm{T}}/\boldsymbol{a}_p,\boldsymbol{c}^{\mathrm{T}})^{\mathrm{T}}$$

其中，$\boldsymbol{a}=(a_1,\cdots,a_p)^{\mathrm{T}}\in\mathbf{R}^p$, $\boldsymbol{b}_h\in\mathbf{R}^{pq}$, $\boldsymbol{c}\in\mathbf{R}^{pq}$，均为列向量。

因为 $\boldsymbol{u}_1=\boldsymbol{f}_1(\overline{\boldsymbol{u}}_0)$，所以由 δ-方法得

$$\sqrt{n}(\boldsymbol{u}_1-\boldsymbol{\mu}_{\boldsymbol{u}_1})\to_d N(\boldsymbol{0},\boldsymbol{\Sigma}_{\boldsymbol{u}_1})$$

其中

$$\boldsymbol{\mu}_{\boldsymbol{u}_1}=\boldsymbol{f}_1(\boldsymbol{\mu}_{\boldsymbol{u}_0})=(p_1,\cdots,p_p,\boldsymbol{m}_1^{\mathrm{T}},\cdots,\boldsymbol{m}_p^{\mathrm{T}},\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{u}_1}=\partial\boldsymbol{f}_1/\partial\boldsymbol{u}^{\mathrm{T}}\big|_{\boldsymbol{\mu}_{\boldsymbol{u}_0}}\boldsymbol{\Sigma}_{\boldsymbol{u}_0}\partial\boldsymbol{f}_1^{\mathrm{T}}/\partial\boldsymbol{u}\big|_{\boldsymbol{\mu}_{\boldsymbol{u}_0}}$$

$$\partial\boldsymbol{f}_1/\partial\boldsymbol{u}^{\mathrm{T}}\big|_{\boldsymbol{\mu}_{\boldsymbol{u}_0}}=\begin{bmatrix}\boldsymbol{I}_p & \boldsymbol{O}_{p,p^2q} & \boldsymbol{O}_{p,pq}\\ \boldsymbol{D}_1^{\mathrm{T}} & \boldsymbol{D}_2^{\mathrm{T}} & \boldsymbol{O}_{p^2q,pq}\\ \boldsymbol{O}_{pq,p} & \boldsymbol{O}_{pq,p^2q} & \boldsymbol{I}_{pq}\end{bmatrix}$$

$\boldsymbol{O}_{p,q}$ 为 $p\times q$ 阶零矩阵，$\boldsymbol{I}_p$ 为 p 阶单位矩阵。

$$\boldsymbol{D}_1=\begin{bmatrix}-\boldsymbol{m}_1^{\mathrm{T}}/p_1 & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & -\boldsymbol{m}_2^{\mathrm{T}}/p_2 & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{0} & \boldsymbol{0} & \cdots & -\boldsymbol{m}_p^{\mathrm{T}}/p_p\end{bmatrix}$$

$$\boldsymbol{D}_2=\begin{bmatrix}\boldsymbol{I}_q/p_1 & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_q/p_2 & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{I}_q/p_p\end{bmatrix}$$

则有

$$\boldsymbol{\Sigma}_{u_1}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{O}_{p,p^2q} & \boldsymbol{B}_{13}\\ \boldsymbol{O}_{p^2q,p} & \boldsymbol{B}_{22}^{*} & \boldsymbol{B}_{23}^{*}\\ \boldsymbol{B}_{13}^{\mathrm{T}} & \boldsymbol{B}_{23}^{*\mathrm{T}} & \boldsymbol{B}_{33}\end{bmatrix}$$

其中

$$\boldsymbol{B}_{22}^{*}=\begin{bmatrix}\boldsymbol{C}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{C}_2 & \cdots & \boldsymbol{0}\\ \vdots & \vdots & \ddots & \vdots\\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{C}_p\end{bmatrix}$$

$$\boldsymbol{C}_h=\tilde{\boldsymbol{V}}_h/p_h^2-\left[(1-p_h)/p_h\right]\boldsymbol{m}_h\boldsymbol{m}_h^{\mathrm{T}}$$

$$\boldsymbol{B}_{23}^{*}=\begin{bmatrix}p_1\boldsymbol{C}_1\\ \vdots\\ p_p\boldsymbol{C}_p\end{bmatrix}$$

第三步，得到 $\hat{\boldsymbol{V}}$ 的渐近分布，因为

$$\boldsymbol{V}=\sum_{h=1}^{p}p_h\sum\left[\boldsymbol{Q}_j\boldsymbol{\Sigma}(\boldsymbol{m}_h-\boldsymbol{\mu})(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{Q}_j^{\mathrm{T}}\right]$$

则$\hat{\boldsymbol{V}}$为相应的样本估计。

定义函数

$$\boldsymbol{f}_2\colon\ \mathbf{R}^{p+p^2q+pq}\ \to\ \mathbf{R}^{q^2}$$

$$\boldsymbol{u}=(\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\boldsymbol{c}^{\mathrm{T}})^{\mathrm{T}}\ \mapsto\ \sum_{h=1}^{p}a_h\sum_{j=1}^{p}\left[\operatorname{vec}\boldsymbol{A}_j\boldsymbol{B}(\boldsymbol{b}_h-\boldsymbol{c})(\boldsymbol{b}_h-\boldsymbol{c})^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}_j^{\mathrm{T}}\right]$$

其中，$\boldsymbol{A}_j\in\mathbf{R}^{q\times pq},\boldsymbol{B}\in\mathbf{R}^{pq\times pq}$，为数量矩阵。

因为

$$\operatorname{vec}(\hat{\boldsymbol{V}})=\boldsymbol{f}_2(\boldsymbol{u}_1),\ \operatorname{vec}(\boldsymbol{V})=\boldsymbol{f}_2(\boldsymbol{\mu}_{\boldsymbol{u}_1})$$

所以

$$\sqrt{n}\left[\operatorname{vec}(\hat{\boldsymbol{V}})-\operatorname{vec}(\boldsymbol{V})\right]\to_d\ \operatorname{vec}(\boldsymbol{U}_{\boldsymbol{V}})$$

其中，$\operatorname{vec}(\boldsymbol{U}_{\boldsymbol{V}})\sim N(0,\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}})$。

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}}&=\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{B}_{11}\boldsymbol{E}_1+\boldsymbol{E}_2^{\mathrm{T}}\boldsymbol{B}_{22}^{*}\boldsymbol{E}_2+\boldsymbol{E}_3^{\mathrm{T}}\boldsymbol{B}_{33}\boldsymbol{E}_3+\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{B}_{13}\boldsymbol{E}_3+\\&\quad\boldsymbol{E}_3^{\mathrm{T}}\boldsymbol{B}_{13}^{\mathrm{T}}\boldsymbol{E}_1+\boldsymbol{E}_2^{\mathrm{T}}\boldsymbol{B}_{23}^{*}\boldsymbol{E}_3+\boldsymbol{E}_3^{\mathrm{T}}\boldsymbol{B}_{23}^{*\mathrm{T}}\boldsymbol{E}_2\end{aligned}$$

$$\boldsymbol{E}_1=\begin{bmatrix}\sum_{j=1}^{p}\left[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\\ \vdots\\ \sum_{j=1}^{p}\left[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\end{bmatrix}$$

$[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]^{[2]}$表示$[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]\otimes[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]$。

$$\boldsymbol{E}_2=\begin{bmatrix}p_1\sum_{j=1}^{p}\left\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\otimes[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]+[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\right\}\\ \vdots\\ p_1\sum_{j=1}^{p}\left\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\otimes[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]+[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\right\}\end{bmatrix}$$

$$\boldsymbol{E}_3=\sum_{h=1}^{p}p_h\sum_{j=1}^{p}\left[(\boldsymbol{I}_{pq}\otimes(\boldsymbol{\mu}-\boldsymbol{m}_h)^{\mathrm{T}}+(\boldsymbol{\mu}-\boldsymbol{m}_h)^{\mathrm{T}}\otimes\boldsymbol{I}_{pq})(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\otimes\boldsymbol{\Sigma}_x^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\right]$$

由于$\boldsymbol{\mu}=\sum_{h=1}^{p}p_h\boldsymbol{m}_h$，所以$\boldsymbol{E}_3=\boldsymbol{0}$，

这样$\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}}$可以化简为$\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}}=\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{B}_{11}\boldsymbol{E}_1+\boldsymbol{E}_2^{\mathrm{T}}\boldsymbol{B}_{22}^{*}\boldsymbol{E}_2$，而

$$
\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{B}_{11}\boldsymbol{E}_1=\begin{bmatrix}\sum_{j=1}^{p}\left[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\\ \vdots\\ \sum_{j=1}^{p}\left[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}p_1(1-p_1) & \cdots & -p_p p_1\\ -p_1p_2 & \cdots & -p_p p_2\\ \vdots & \ddots & \vdots\\ -p_1p_p & \cdots & p_p(1-p_p)\end{bmatrix}
$$

$$
\begin{bmatrix}\sum_{j=1}^{p}\left[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\\ \vdots\\ \sum_{j=1}^{p}\left[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\end{bmatrix}
$$

$$
=\begin{bmatrix}p_1\left(\sum_{j=1}^{p}\left[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}-\sum_{h=1}^{p}p_h\sum_{j=1}^{p}\left[\left(\boldsymbol{m}_h-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\right)\\ \vdots\\ p_p\left(\sum_{j}^{p}\left[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}-\sum_{h=1}^{p}p_h\sum_{j=1}^{p}\left[\left(\boldsymbol{m}_h-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\right)\end{bmatrix}^{\mathrm{T}}
$$

$$
\begin{bmatrix}\sum_{j=1}^{p}\left[(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\\ \vdots\\ \sum_{j=1}^{p}\left[(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\end{bmatrix}
$$

$$
=\sum_{h=1}^{p}p_h\left(\sum_{j=1}^{p}\left[\left(\boldsymbol{m}_h-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\right)^{\mathrm{T}}\left(\sum_{j=1}^{p}\left[\left(\boldsymbol{m}_h-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\right)-
$$

$$
\left(\sum_{h=1}^{p}p_h\sum_{j=1}^{p}\left[\left(\boldsymbol{m}_h-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\right)^{\mathrm{T}}\left(\sum_{h=1}^{p}p_h\sum_{j=1}^{p}\left[\left(\boldsymbol{m}_h-\boldsymbol{\mu}\right)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\right]^{[2]}\right)
$$

$$
=\mathrm{Cov}\left(\sum_{j=1}^{p}\left[\boldsymbol{Q}_j\boldsymbol{\Sigma}^{-1}\boldsymbol{E}\left(\boldsymbol{Z}\,|\,\boldsymbol{Y}\right)\right]^{[2]}\right)
$$

$$
\boldsymbol{E}_2^{\mathrm{T}}\boldsymbol{B}_{22}^{*}\boldsymbol{E}_2=\sum_{h=1}^{p}\left\{p_h\sum_{j=1}^{p}\left((\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\otimes[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]+[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}]\otimes\right.\right.
$$

$$
\left.\left.(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})\right)^{\mathrm{T}}\boldsymbol{C}_h\right.
$$

$$p_h\sum_{j=1}^{p}\Big((\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\otimes[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]+[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\Big)\Bigg\}$$

$$=\sum_{h=1}^{p}\Bigg\{\sum_{j=1}^{p}\Big((\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\otimes[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]+[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\Big)^{\mathrm T}$$

$$[\boldsymbol{E}(\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm T}\delta_h)/p_h-\boldsymbol{m}_h\boldsymbol{m}_h^{\mathrm T}]$$

$$\sum_{j=1}^{p}\Big\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\otimes[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]+[(\boldsymbol{m}_h-\boldsymbol{\mu})^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\Big\}$$

$$=\boldsymbol{E}\Big[\boldsymbol{K}(\boldsymbol{Y})^{\mathrm T}\mathrm{Cov}(\boldsymbol{Z}|\ Y)\boldsymbol{K}(\boldsymbol{Y})\Big]$$

其中

$$\boldsymbol{K}(\boldsymbol{Y})=\sum_{j=1}^{p}\Big\{(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\otimes[\boldsymbol{E}(\boldsymbol{Z}|\ Y)^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]+[\boldsymbol{E}(\boldsymbol{Z}|\ Y)^{\mathrm T}\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}]\otimes(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})\Big\}$$

得到协方差阵式（4.9）。

证毕。

定理 4.3.2 在广义多项选择模型中，若 $\mathrm{vec}(\boldsymbol{X})$ 存在有限的四阶矩，当 $\boldsymbol{\Sigma}$ 由数据估计得到时，有

$$\sqrt{n}(\hat{\boldsymbol{V}}-\boldsymbol{V})\to_d \boldsymbol{U}_{\boldsymbol{V}}^*$$

其中，$\mathrm{vec}(\boldsymbol{U}_{\boldsymbol{V}}^*)\sim N(\boldsymbol{0},\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}^*})$。

$$\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}^*}=[\boldsymbol{J}_1^{\mathrm T}\quad \boldsymbol{J}_2^{\mathrm T}]\begin{bmatrix}\boldsymbol{F}_{11} & -\boldsymbol{F}_{12}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\\ -(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{F}_{12}^{\mathrm T} & (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{F}_{22}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\end{bmatrix}\begin{bmatrix}\boldsymbol{J}_1\\ \boldsymbol{J}_2\end{bmatrix}\quad(4.10)$$

其中

$$\boldsymbol{J}_1=\sum_{j=1}^{p}\boldsymbol{I}_{(pq)^2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm T})$$

$$\boldsymbol{J}_2=\sum_{j=1}^{p}\Big\{\Big[\boldsymbol{I}_{(pq)^2}\otimes(\boldsymbol{W}\boldsymbol{\Sigma}^{-1})+(\boldsymbol{\Sigma}^{-1}\boldsymbol{W})\otimes\boldsymbol{I}_{(pq)^2}\Big]\Big[\boldsymbol{Q}_j^{\mathrm T}\otimes\boldsymbol{Q}_j^{\mathrm T}\Big]\Big\}$$

$$\boldsymbol{F}_{11}=\boldsymbol{R}_1^{\mathrm T}\boldsymbol{B}_{11}\boldsymbol{R}_1+\boldsymbol{R}_2^{\mathrm T}\boldsymbol{B}_{22}^*\boldsymbol{R}_2+\boldsymbol{R}_3^{\mathrm T}\boldsymbol{B}_{33}\boldsymbol{R}_3+\boldsymbol{R}_1^{\mathrm T}\boldsymbol{B}_{13}\boldsymbol{R}_3+\boldsymbol{R}_3^{\mathrm T}\boldsymbol{B}_{13}^{\mathrm T}\boldsymbol{R}_1+\boldsymbol{R}_2^{\mathrm T}\boldsymbol{B}_{23}^*\boldsymbol{R}_3+\boldsymbol{R}_3^{\mathrm T}\boldsymbol{B}_{23}^{*\mathrm T}\boldsymbol{R}_2$$

$$\boldsymbol{F}_{12}=\boldsymbol{R}_1^{\mathrm T}\boldsymbol{B}_{13}\boldsymbol{R}_4+\boldsymbol{R}_1^{\mathrm T}\boldsymbol{B}_{14}+\boldsymbol{R}_2^{\mathrm T}\boldsymbol{B}_{23}^*\boldsymbol{R}_4+\boldsymbol{R}_2^{\mathrm T}\boldsymbol{B}_{24}^*+\boldsymbol{R}_3^{\mathrm T}\boldsymbol{B}_{33}\boldsymbol{R}_4+\boldsymbol{R}_3^{\mathrm T}\boldsymbol{B}_{34}$$

$$\boldsymbol{F}_{22}=\boldsymbol{R}_4^{\mathrm T}\boldsymbol{B}_{33}\boldsymbol{R}_4+\boldsymbol{R}_4^{\mathrm T}\boldsymbol{B}_{34}+\boldsymbol{B}_{34}^{\mathrm T}\boldsymbol{R}_4+\boldsymbol{B}_{44}$$

$\boldsymbol{B}_{11},\boldsymbol{B}_{13},\boldsymbol{B}_{33},\boldsymbol{B}_{22}^*,\boldsymbol{B}_{23}^*$ 与定理 4.3.1 中相同，有

$$\boldsymbol{R}_1=\begin{bmatrix}(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm T}\otimes(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm T}\\ \vdots\\ (\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm T}\otimes(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm T}\end{bmatrix}$$

$$\boldsymbol{R}_2=\begin{bmatrix} p_1(\boldsymbol{I}_{pq}\otimes\boldsymbol{m}_1^{\mathrm{T}}+\boldsymbol{m}_1^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}}) \\ \vdots \\ p_p(\boldsymbol{I}_{pq}\otimes\boldsymbol{m}_p^{\mathrm{T}}+\boldsymbol{m}_p^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}}) \end{bmatrix}$$

$$\boldsymbol{R}_3=\sum_{h=1}^{p}p_h(\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}}+\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{I}_{pq}\otimes\boldsymbol{m}_h^{\mathrm{T}}-\boldsymbol{m}_h^{\mathrm{T}}\otimes\boldsymbol{I}_{pq})$$

$$\boldsymbol{R}_4=(\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}})+(\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq})$$

$$\boldsymbol{B}_{14}=\begin{bmatrix} \boldsymbol{\nu}_1-p_1\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}} \\ \vdots \\ \boldsymbol{\nu}_p-p_p\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}} \end{bmatrix}$$

$$\boldsymbol{B}_{24}=\begin{bmatrix} \tilde{\boldsymbol{G}}_1-\tilde{\boldsymbol{m}}_1\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}} \\ \vdots \\ \tilde{\boldsymbol{G}}_p-\tilde{\boldsymbol{m}}_p\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}} \end{bmatrix}$$

$$\boldsymbol{B}_{34}=[\boldsymbol{G}-\boldsymbol{\mu}\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}}]$$

$$\boldsymbol{B}_{44}=[\boldsymbol{H}-\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^{\mathrm{T}})^{\mathrm{T}}]$$

$$\boldsymbol{B}_{24}^{*}=\begin{bmatrix} \tilde{\boldsymbol{G}}_1/p_1-\tilde{\boldsymbol{m}}_1\boldsymbol{\nu}_1^{\mathrm{T}}/p_1^2 \\ \vdots \\ \tilde{\boldsymbol{G}}_p/p_p-\tilde{\boldsymbol{m}}_1\boldsymbol{\nu}_p^{\mathrm{T}}/p_p^2 \end{bmatrix}$$

$$\boldsymbol{\nu}_h=\boldsymbol{E}\{[\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\otimes\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\delta_h\}$$

$$\tilde{\boldsymbol{G}}_h=\boldsymbol{E}\{\mathrm{vec}(\boldsymbol{X})[\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\otimes\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\delta_h\}$$

$$\boldsymbol{G}=\boldsymbol{E}\{\mathrm{vec}(\boldsymbol{X})[\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\otimes\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\}$$

$$\boldsymbol{H}=\boldsymbol{E}\{[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\otimes[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\}$$

证明：第一步，将$\hat{\boldsymbol{V}}$写成等价形式为

$$\hat{\boldsymbol{V}}=\sum_{j=1}^{p}\left\{\boldsymbol{Q}_j\hat{\boldsymbol{\Sigma}}_x\left[\sum_{h=1}^{p}p_h(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}})(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}})^{\mathrm{T}}\right]\hat{\boldsymbol{\Sigma}}\boldsymbol{Q}_j^{\mathrm{T}}\right\}$$

令

$$\boldsymbol{U}^{(i)}=\begin{bmatrix} \boldsymbol{u}_0^{(i)} \\ \mathrm{vec}[\mathrm{vec}(\boldsymbol{X}^{(i)})\mathrm{vec}(\boldsymbol{X}^{(i)})^{\mathrm{T}}] \end{bmatrix}$$

则有

$$\boldsymbol{U}^{(i)}=\begin{bmatrix} \boldsymbol{u}_0^{(i)} \\ \mathrm{vec}[\mathrm{vec}(\boldsymbol{X}^{(i)})\mathrm{vec}(\boldsymbol{X}^{(i)})^{\mathrm{T}}] \end{bmatrix}$$

$$\boldsymbol{\Sigma}_U=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{B}_{13} & \boldsymbol{B}_{14}\\ \boldsymbol{B}_{12}^{\mathrm{T}} & \boldsymbol{B}_{22} & \boldsymbol{B}_{23} & \boldsymbol{B}_{24}\\ \boldsymbol{B}_{13}^{\mathrm{T}} & \boldsymbol{B}_{23}^{\mathrm{T}} & \boldsymbol{B}_{33} & \boldsymbol{B}_{34}\\ \boldsymbol{B}_{14}^{\mathrm{T}} & \boldsymbol{B}_{24}^{\mathrm{T}} & \boldsymbol{B}_{34}^{\mathrm{T}} & \boldsymbol{B}_{44}\end{bmatrix}$$

其中，$\boldsymbol{B}_{11}\cdots\boldsymbol{B}_{33}$ 为上述定理中的形式。

$$\boldsymbol{B}_{14}=\begin{bmatrix}\boldsymbol{\nu}_1-p_1\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}\\ \vdots\\ \boldsymbol{\nu}_p-p_p\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}\end{bmatrix}$$

$$\boldsymbol{B}_{24}=\begin{bmatrix}\tilde{\boldsymbol{G}}_1-\tilde{\boldsymbol{m}}_1\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}\\ \vdots\\ \tilde{\boldsymbol{G}}_p-\tilde{\boldsymbol{m}}_p\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}\end{bmatrix}$$

$$\boldsymbol{B}_{34}=[\boldsymbol{G}-\boldsymbol{\mu}\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}]$$

$$\boldsymbol{B}_{44}=[\boldsymbol{H}-\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}]$$

$$\boldsymbol{\nu}_h=\boldsymbol{E}\{\mathrm{vec}[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\delta_h\}=\boldsymbol{E}\{[\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\otimes\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\delta_h\}$$

$$\tilde{\boldsymbol{G}}_h=\boldsymbol{E}\{\mathrm{vec}(\boldsymbol{X})\mathrm{vec}[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]^{\mathrm{T}}\delta_h\}=\boldsymbol{E}\{\mathrm{vec}(\boldsymbol{X})[\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\otimes\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\delta_h\}$$

$$\boldsymbol{G}=\boldsymbol{E}\{\mathrm{vec}(\boldsymbol{X})\mathrm{vec}[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]^{\mathrm{T}}\}=\boldsymbol{E}\{\mathrm{vec}(\boldsymbol{X})[\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}\otimes\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\}$$

$$\boldsymbol{H}=\boldsymbol{E}\{[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\otimes[\mathrm{vec}(\boldsymbol{X})\mathrm{vec}(\boldsymbol{X})^{\mathrm{T}}]\}$$

令

$$\bar{\boldsymbol{U}}=\frac{1}{n}\sum_{i=1}^{p}\boldsymbol{U}^{(i)}$$

由中心有限定理得

$$\sqrt{n}(\bar{\boldsymbol{U}}-\boldsymbol{\mu}_U)\rightarrow_d N(\boldsymbol{0},\boldsymbol{\Sigma}_U)$$

第二步，得到构成 $\hat{\boldsymbol{V}}$ 和 $\hat{\boldsymbol{\Sigma}}_x$ 的随机向量的分布，令

$$\boldsymbol{U}_1=\begin{bmatrix}\boldsymbol{u}_1\\ \dfrac{1}{n}\sum_{i=1}^{n}\mathrm{vec}\left[\mathrm{vec}(\boldsymbol{X}^{(i)}\mathrm{vec}(\boldsymbol{X}^{(i)})^{\mathrm{T}}\right]\end{bmatrix}$$

定义函数：

$$\begin{array}{llll}\boldsymbol{g}_2: & \mathbf{R}^{p+p^2q+pq+(pq)^2} & \rightarrow & \mathbf{R}^{p+p^2q+pq+(pq)^2}\\ & (\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\boldsymbol{c}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}}) & \mapsto & (\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}}/\boldsymbol{a}_1,\cdots,\boldsymbol{b}_p^{\mathrm{T}}/\boldsymbol{a}_p,\boldsymbol{c}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}})\end{array}$$

其中，$\boldsymbol{a}=(a_1,\cdots,a_p)^{\mathrm{T}}\in\mathbf{R}^p,\boldsymbol{b}_h\in\mathbf{R}^{pq},\boldsymbol{c}\in\mathbf{R}^{pq},\boldsymbol{d}\in\mathbf{R}^{(pq)^2}$ 均为列向量。

因为

$$\boldsymbol{U}_1=\boldsymbol{g}_1(\hat{\boldsymbol{U}})$$

$$\boldsymbol{\mu}_{U_1}=\boldsymbol{g}_1(\boldsymbol{\mu}_U)=\left[p_1,\cdots,p_p,\boldsymbol{m}_1^{\mathrm{T}},\cdots,\boldsymbol{m}_p^{\mathrm{T}},\boldsymbol{\mu}^{\mathrm{T}},\mathrm{vec}(\boldsymbol{\Sigma}+\boldsymbol{\mu\mu}^{\mathrm{T}})^{\mathrm{T}}\right]^{\mathrm{T}}$$

由 δ-方法得

$$\sqrt{n}(\boldsymbol{U}_1-\boldsymbol{\mu}_{U_1})\to_d N(\boldsymbol{0},\boldsymbol{\Sigma}_{U_1})$$

其中

$$\boldsymbol{\Sigma}_{U_1}=\begin{bmatrix}\boldsymbol{B}_{11} & \boldsymbol{O}_{p,p^2q} & \boldsymbol{B}_{13} & \boldsymbol{B}_{14}\\ \boldsymbol{O}_{p^2q,pq} & \boldsymbol{B}_{22}^{*} & \boldsymbol{B}_{23}^{*} & \boldsymbol{B}_{24}^{*}\\ \boldsymbol{B}_{13}^{\mathrm{T}} & \boldsymbol{B}_{23}^{*\mathrm{T}} & \boldsymbol{B}_{33} & \boldsymbol{B}_{34}\\ \boldsymbol{B}_{14}^{\mathrm{T}} & \boldsymbol{B}_{24}^{*\mathrm{T}} & \boldsymbol{B}_{34}^{\mathrm{T}} & \boldsymbol{B}_{44}\end{bmatrix}$$

$\boldsymbol{B}_{22}^{*},\boldsymbol{B}_{23}^{*}$ 与定理 4.3.1 中相同，有

$$\boldsymbol{B}_{24}^{*}=\begin{bmatrix}\tilde{\boldsymbol{G}}_1/p_1-\tilde{\boldsymbol{m}}_1\boldsymbol{\nu}_1^{\mathrm{T}}/p_1^2\\ \vdots\\ \tilde{\boldsymbol{G}}_p/p_p-\tilde{\boldsymbol{m}}_1\boldsymbol{\nu}_p^{\mathrm{T}}/p_p^2\end{bmatrix}$$

第三步，记

$$\hat{\boldsymbol{W}}=\sum_{h=1}^{p}\hat{p}_h(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}})(\hat{\boldsymbol{m}}_h-\hat{\boldsymbol{\mu}})^{\mathrm{T}}$$

得到

$$\sqrt{n}\left(\begin{bmatrix}\mathrm{vec}(\hat{\boldsymbol{W}})\\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}_x)\end{bmatrix}-\begin{bmatrix}\mathrm{vec}(\boldsymbol{W})\\ \mathrm{vec}(\boldsymbol{\Sigma})\end{bmatrix}\right)$$

的渐近分布。

定义函数：

$$\begin{aligned}\boldsymbol{g}_2:\ \mathbf{R}^{p+p^2q+pq+(pq)^2} &\to \mathbf{R}^{2(pq)^2}\\ (\boldsymbol{a}^{\mathrm{T}},\boldsymbol{b}_1^{\mathrm{T}},\cdots,\boldsymbol{b}_p^{\mathrm{T}},\boldsymbol{c}^{\mathrm{T}},\boldsymbol{d}^{\mathrm{T}})^{\mathrm{T}} &\mapsto \begin{pmatrix}\sum_{h=1}^{p}a_h\mathrm{vec}[(\boldsymbol{b}_h-\boldsymbol{c})(\boldsymbol{b}_h-\boldsymbol{c})^{\mathrm{T}}]\\ \mathrm{vec}(\boldsymbol{d})-\mathrm{vec}(\boldsymbol{cc}^{\mathrm{T}})\end{pmatrix}\end{aligned}$$

因为

$$\boldsymbol{g}_2(\boldsymbol{U}_1)=\begin{bmatrix}\mathrm{vec}(\hat{\boldsymbol{W}})\\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}})\end{bmatrix}$$

$$\boldsymbol{g}_2(\boldsymbol{\mu}_{U_1})=\begin{bmatrix}\mathrm{vec}(\boldsymbol{W})\\ \mathrm{vec}(\boldsymbol{\Sigma})\end{bmatrix}$$

所以由 δ-方法得

$$\sqrt{n}\left(\begin{bmatrix}\text{vec}(\hat{\boldsymbol{W}})\\ \text{vec}(\hat{\boldsymbol{\Sigma}})\end{bmatrix}-\begin{bmatrix}\text{vec}(\boldsymbol{W})\\ \text{vec}(\boldsymbol{\Sigma})\end{bmatrix}\right)\to_d N(\boldsymbol{0},\ \boldsymbol{\Sigma}_{U_2})$$

其中

$$\boldsymbol{\Sigma}_{U_2}=\begin{bmatrix}\boldsymbol{F}_{11} & \boldsymbol{F}_{12}\\ \boldsymbol{F}_{12}^{\mathrm{T}} & \boldsymbol{F}_{22}\end{bmatrix}$$

$$\boldsymbol{F}_{11}=\boldsymbol{R}_1^{\mathrm{T}}\boldsymbol{B}_{11}\boldsymbol{R}_1+\boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{B}_{22}^{*}\boldsymbol{R}_2+\boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{B}_{33}\boldsymbol{R}_3+\boldsymbol{R}_1^{\mathrm{T}}\boldsymbol{B}_{13}\boldsymbol{R}_3+\boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{B}_{13}^{\mathrm{T}}\boldsymbol{R}_1+\boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{B}_{23}^{*}\boldsymbol{R}_3+\boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{B}_{23}^{*\mathrm{T}}\boldsymbol{R}_2$$

$$\boldsymbol{F}_{12}=\boldsymbol{R}_1^{\mathrm{T}}\boldsymbol{B}_{13}\boldsymbol{R}_4+\boldsymbol{R}_1^{\mathrm{T}}\boldsymbol{B}_{14}+\boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{B}_{23}^{*}\boldsymbol{R}_4+\boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{B}_{24}^{*}+\boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{B}_{33}\boldsymbol{R}_4+\boldsymbol{R}_3^{\mathrm{T}}\boldsymbol{B}_{34}$$

$$\boldsymbol{F}_{22}=\boldsymbol{R}_4^{\mathrm{T}}\boldsymbol{B}_{33}\boldsymbol{R}_4+\boldsymbol{R}_4^{\mathrm{T}}\boldsymbol{B}_{34}+\boldsymbol{B}_{34}^{\mathrm{T}}\boldsymbol{R}_4+\boldsymbol{B}_{44}$$

$$\boldsymbol{R}_1=\begin{bmatrix}(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\otimes(\boldsymbol{m}_1-\boldsymbol{\mu})^{\mathrm{T}}\\ \vdots\\ (\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\otimes(\boldsymbol{m}_p-\boldsymbol{\mu})^{\mathrm{T}}\end{bmatrix}$$

$$\boldsymbol{R}_2=\begin{bmatrix}p_1(\boldsymbol{I}_{pq}\otimes\boldsymbol{m}_1^{\mathrm{T}}+\boldsymbol{m}_1^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}})\\ \vdots\\ p_p(\boldsymbol{I}_{pq}\otimes\boldsymbol{m}_p^{\mathrm{T}}+\boldsymbol{m}_p^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}})\end{bmatrix}$$

$$\boldsymbol{R}_3=\sum_{h=1}^{p}p_h(\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}}+\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq}-\boldsymbol{I}_{pq}\otimes\boldsymbol{m}_h^{\mathrm{T}}-\boldsymbol{m}_h^{\mathrm{T}}\otimes\boldsymbol{I}_{pq})$$

$$\boldsymbol{R}_4=(\boldsymbol{I}_{pq}\otimes\boldsymbol{\mu}^{\mathrm{T}})+(\boldsymbol{\mu}^{\mathrm{T}}\otimes\boldsymbol{I}_{pq})$$

第四步，得到$\sqrt{n}\left[\text{vec}(\hat{\boldsymbol{V}})-\text{vec}(\boldsymbol{V})\right]$的渐近分布，利用一阶近似

$$\hat{\boldsymbol{\Sigma}}^{-1}\approx\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}$$

得到

$$\sqrt{n}\left(\begin{bmatrix}\text{vec}(\hat{\boldsymbol{W}})\\ \text{vec}(\hat{\boldsymbol{\Sigma}}^{-1})\end{bmatrix}-\begin{bmatrix}\text{vec}(\boldsymbol{W})\\ \text{vec}(\boldsymbol{\Sigma}^{-1})\end{bmatrix}\right)\approx\boldsymbol{S}\sqrt{n}\left(\begin{bmatrix}\text{vec}(\hat{\boldsymbol{W}})\\ \text{vec}(\hat{\boldsymbol{\Sigma}})\end{bmatrix}-\begin{bmatrix}\text{vec}(\boldsymbol{W})\\ \text{vec}(\boldsymbol{\Sigma})\end{bmatrix}\right)$$

其中

$$\boldsymbol{S}=\begin{bmatrix}\boldsymbol{I}_{(pq)^2} & \boldsymbol{O}_{(pq)^2}\\ \boldsymbol{O}_{(pq)^2} & -(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\end{bmatrix}$$

因此得到

$$\sqrt{n}\left(\begin{bmatrix}\text{vec}(\hat{\boldsymbol{W}})\\ \text{vec}(\hat{\boldsymbol{\Sigma}})\end{bmatrix}-\begin{bmatrix}\text{vec}(\boldsymbol{W})\\ \text{vec}(\boldsymbol{\Sigma})\end{bmatrix}\right)\to_d N(\boldsymbol{0},\ \boldsymbol{\Sigma}_{U_3})$$

其中

$$\boldsymbol{\Sigma}_{U_3}=\begin{bmatrix}\boldsymbol{F}_{11} & -\boldsymbol{F}_{12}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\\ -(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{F}_{12}^{\mathrm{T}} & (\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\boldsymbol{F}_{22}(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1})\end{bmatrix}$$

定义函数：

$$\boldsymbol{g}_3\colon \mathbf{R}^{2(pq)^2}\ \to\ \mathbf{R}^{q^2}$$

$$\begin{pmatrix}\mathrm{vec}(\boldsymbol{A})\\ \mathrm{vec}(\boldsymbol{B})\end{pmatrix}\ \mapsto\ \sum_{j=1}^{p}\mathrm{vec}(\boldsymbol{C}_j\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C}_j^{\mathrm{T}})$$

其中，$\boldsymbol{A},\boldsymbol{B}\in\mathbf{R}^{pq\times pq},\boldsymbol{C}_j\in\mathbf{R}^{q\times pq}$，为数量矩阵，因为

$$\mathrm{vec}(\hat{\boldsymbol{V}})=\boldsymbol{g}_3\begin{pmatrix}\mathrm{vec}(\hat{\boldsymbol{W}})\\ \mathrm{vec}(\hat{\boldsymbol{\Sigma}}^{-1})\end{pmatrix}$$

$$\mathrm{vec}(\boldsymbol{V})=\boldsymbol{g}_3\begin{pmatrix}\mathrm{vec}(\boldsymbol{W})\\ \mathrm{vec}(\boldsymbol{\Sigma}^{-1})\end{pmatrix}$$

得

$$\sqrt{n}[\mathrm{vec}(\hat{\boldsymbol{V}})-\mathrm{vec}(\boldsymbol{V})]\to_d N(\boldsymbol{0},\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}^*})$$

$$\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}^*}=[\boldsymbol{J}_1^{\mathrm{T}}\quad \boldsymbol{J}_2^{\mathrm{T}}]\boldsymbol{\Sigma}_{\boldsymbol{U}_3}\begin{bmatrix}\boldsymbol{J}_1\\ \boldsymbol{J}_2\end{bmatrix}$$

其中

$$\boldsymbol{J}_1=\sum_{j=1}^{p}\boldsymbol{I}_{(pq)^2}(\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}}\otimes\boldsymbol{\Sigma}^{-1}\boldsymbol{Q}_j^{\mathrm{T}})$$

$$\boldsymbol{J}_2=\sum_{j=1}^{p}\left\{\left[\boldsymbol{I}_{(pq)^2}\otimes(\boldsymbol{W}\boldsymbol{\Sigma}^{-1})+(\boldsymbol{\Sigma}^{-1}\boldsymbol{W})\otimes\boldsymbol{I}_{(pq)^2}\right]\left[\boldsymbol{Q}_j^{\mathrm{T}}\otimes\boldsymbol{Q}_j^{\mathrm{T}}\right]\right\}$$

证毕。

引理 4.3.1（Tyler[67]引理 4.1）　令 $\boldsymbol{M}$ 为一个 $p\times p$ 矩阵，$\boldsymbol{M}$ 关于一个正定矩阵 $\boldsymbol{\Gamma}$ 是对称的，也就是说 $\boldsymbol{\Gamma M}$ 是一个对称矩阵。令 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$ 为 $\boldsymbol{M}$ 的特征值，ω 是 $\{\lambda_1,\lambda_2,\cdots,\lambda_p\}$ 的一个子集，$\omega=\{\lambda_i,\cdots,\lambda_{i+m}\}$, $i+m\leqslant p$ 并且满足 $\lambda_{i-1}\neq\lambda_i$，$\lambda_{i+m-1}\neq\lambda_{i+m}$。$\boldsymbol{M}_n$ 是 $\boldsymbol{M}$ 的一个估计序列，使得 $a_n(\boldsymbol{M}_n-\boldsymbol{M})$ 依概率收敛到一个正态分布，其中 a_n 是一个递增的实数列。

令 $\boldsymbol{P}_\lambda$ 表示矩阵 $\boldsymbol{M}$ 对应特征值 λ 的特征投影，即 $\boldsymbol{P}_\lambda=\gamma\gamma^{\mathrm{T}}\boldsymbol{\Gamma}$，其中 γ 是特征值 λ 对应的特征向量。记 $\boldsymbol{P}_0=\boldsymbol{\Sigma}_{\lambda\in\omega}\boldsymbol{P}_\lambda$，$\hat{\boldsymbol{P}}_0=\boldsymbol{\Sigma}_{\lambda\in\omega}\hat{\boldsymbol{P}}_\lambda$，其中 $\hat{\boldsymbol{P}}_\lambda$ 为估计矩阵 $\boldsymbol{M}_n$ 的特征值对应的特征投影。

令 $d_0=\min\{\lambda_{i-1}-\lambda_i,\lambda_{i+m-1}-\lambda_{i+m}\}$，$d_1=\lambda_i-\lambda_{i+m}$。同时定义矩阵 $\boldsymbol{M}$ 的模为

$$\|\boldsymbol{M}\|=[\max\ \mathrm{eigenvalue}(\boldsymbol{\Gamma}^{-1}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{M})]^{1/2}$$

如果有$\|\boldsymbol{M}_n-\boldsymbol{M}\|\leqslant d_0/2$，则

$$\hat{\boldsymbol{P}}_0=\boldsymbol{P}_0-\sum_{i=1}^{m}\left[\boldsymbol{P}_{\lambda_i}(\boldsymbol{M}_n-\boldsymbol{M})(\boldsymbol{M}-\lambda_i\boldsymbol{I})^{+}+(\boldsymbol{M}-\lambda_i\boldsymbol{I})^{+}(\boldsymbol{M}_n-\boldsymbol{M})\boldsymbol{P}_{\lambda_i}\right]+\boldsymbol{R}_n$$

式中，$(\boldsymbol{M}-\lambda_i\boldsymbol{I})^{+}$表示$(\boldsymbol{M}-\lambda_i\boldsymbol{I})$的 Moore-Penrose 广义逆，并且$\boldsymbol{R}_n$满足

$$\boldsymbol{R}_n\leqslant(1+d_1/d_0)(2\|\boldsymbol{M}_n-\boldsymbol{M}\|/d_0)^2(1-2\|\boldsymbol{M}_n-\boldsymbol{M}\|/d_0)^{-1}$$

引理 4.3.1 是 Tyler 文中的引理 4.1，此处略去证明。

定理 4.3.3 令λ、$\boldsymbol{\beta}$是式（4.6）中矩阵$\boldsymbol{V}$的特征值及其对应的特征向量，$\hat{\lambda}$、$\hat{\boldsymbol{\beta}}$是估计矩阵$\hat{\boldsymbol{V}}$的最大特征值及其对应的特征向量。在定理4.31或定理4.3.2的假定下，若$\boldsymbol{\beta}$和$\hat{\boldsymbol{\beta}}$满足式（4.7）和式（4.8），则样本估计$\hat{\boldsymbol{\beta}}$的渐近分布为

$$\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\to_d\frac{1}{\lambda}(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\boldsymbol{U}_{\boldsymbol{V}}\boldsymbol{\beta}\qquad(4.11)$$

式中，$\frac{1}{\lambda}(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\boldsymbol{U}_{\boldsymbol{V}}\boldsymbol{\beta}$服从多元正态分布，期望为 **0**，协方差阵为

$$\frac{1}{\lambda^2}\left[\boldsymbol{\beta}^{\mathrm{T}}\otimes(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\right]\boldsymbol{\Sigma}_{\boldsymbol{U}_{\boldsymbol{V}}}\left[\boldsymbol{\beta}\otimes(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\right]\qquad(4.12)$$

证明：定义$\boldsymbol{V}$的模为

$$\|\boldsymbol{V}\|=[\max\ \mathrm{eigenvalue}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})]^{\frac{1}{2}}$$

因为$\hat{\boldsymbol{V}}$收敛到$\boldsymbol{V}$，所以当n充分大时有$\|\hat{\boldsymbol{V}}-\boldsymbol{V}\|\leqslant\frac{\lambda}{2}$，由引理 4.3.1 得

$$\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathrm{T}}=\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}-[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}(\hat{\boldsymbol{V}}-\boldsymbol{V})(\boldsymbol{V}-\lambda\boldsymbol{I}_q)^{+}+(\boldsymbol{V}-\lambda\boldsymbol{I}_q)^{+}(\hat{\boldsymbol{V}}-\boldsymbol{V})\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}]+\boldsymbol{R}_n$$

其中

$$\boldsymbol{R}_n\leqslant(\|\hat{\boldsymbol{V}}-\boldsymbol{V}\|/\lambda)^2(1-2\|\hat{\boldsymbol{V}}-\boldsymbol{V}\|/\lambda)^{-1}$$

$(\hat{\boldsymbol{V}}-\lambda\boldsymbol{I}_q)^{+}$表示$(\hat{\boldsymbol{V}}-\lambda\boldsymbol{I}_q)$的 Moore-Penrose 广义逆。

因为$\hat{\boldsymbol{V}}-\boldsymbol{V}=O\sqrt{n}$，所以有

$$\sqrt{n}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathrm{T}}-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\to_d-[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{V}}(\boldsymbol{V}-\lambda\boldsymbol{I}_q)^{+}+(\boldsymbol{V}-\lambda\boldsymbol{I}_q)^{+}\boldsymbol{U}_{\boldsymbol{V}}\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}]$$

又因为

$$(\boldsymbol{V}-\lambda\boldsymbol{I}_q)^{+}=-\frac{1}{\lambda}(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})^{+}=-\frac{1}{\lambda}(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})$$

得

$$\sqrt{n}(\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}^{\mathrm{T}}-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\to_d\frac{1}{\lambda}[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{U}_{\boldsymbol{V}}(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})+(\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}})\boldsymbol{U}_{\boldsymbol{V}}\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}]\qquad(4.13)$$

将式（4.13）两端都乘以$\boldsymbol{\beta}$，右端为$\frac{1}{\lambda}[\boldsymbol{I}_q-\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}]\boldsymbol{U}_{\boldsymbol{V}}\boldsymbol{\beta}$，左端为

$$\sqrt{n}(\hat{\boldsymbol\beta}\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta-\boldsymbol\beta)=\sqrt{n}(\hat{\boldsymbol\beta}-\boldsymbol\beta)+\hat{\boldsymbol\beta}\sqrt{n}(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta-1)$$

由于

$$\begin{aligned}&\mathrm{tr}[\sqrt{n}(\hat{\boldsymbol\beta}\hat{\boldsymbol\beta}^{\mathrm T}-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})\boldsymbol\beta\boldsymbol\beta^{\mathrm T}]\\&=\sqrt{n}(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta)^2-1)\\&=\sqrt{n}(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta-1)(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta+1)\end{aligned}$$

所以有

$$\sqrt{n}(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta-1)=\frac{\mathrm{tr}[\sqrt{n}(\hat{\boldsymbol\beta}\hat{\boldsymbol\beta}^{\mathrm T}-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})\boldsymbol\beta\boldsymbol\beta^{\mathrm T}]}{(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta+1)} \tag{4.14}$$

再根据式（4.13）得

$$\hat{\boldsymbol\beta}\hat{\boldsymbol\beta}^{\mathrm T}\to_d\boldsymbol\beta\boldsymbol\beta^{\mathrm T}$$

比较两个矩阵的第一行可得

$$\hat\beta_1\hat\beta_j\to_d\beta_1\beta_j,\ j=1,\cdots,q$$

由 $\hat\beta_1\geqslant0,\beta_1>0$，得 $\hat\beta_1\to_d\beta_1$，进一步得到 $\hat\beta_j\to_d\beta_j$，所以有 $\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta\to_d1$。

又因为

$$\begin{aligned}&\mathrm{tr}\left[\sqrt{n}(\hat{\boldsymbol\beta}\hat{\boldsymbol\beta}^{\mathrm T}-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})\boldsymbol\beta\boldsymbol\beta^{\mathrm T}\right]\to_d\mathrm{tr}\left\{\frac{1}{\lambda}[\boldsymbol\beta\boldsymbol\beta^{\mathrm T}\boldsymbol U_{\boldsymbol V}(\boldsymbol I_q-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})+\right.\\&\left.(\boldsymbol I_q-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})\boldsymbol U_{\boldsymbol V}\boldsymbol\beta\boldsymbol\beta^{\mathrm T}]\boldsymbol\beta\boldsymbol\beta^{\mathrm T}\right\}=0\end{aligned} \tag{4.15}$$

由式（4.14）和式（4.15）得到

$$\sqrt{n}(\hat{\boldsymbol\beta}^{\mathrm T}\boldsymbol\beta-1)\to_d0$$

所以

$$\sqrt{n}(\hat{\boldsymbol\beta}-\boldsymbol\beta)\to_d\frac{1}{\lambda}[\boldsymbol I_q-\boldsymbol\beta\boldsymbol\beta^{\mathrm T}]\boldsymbol U_{\boldsymbol V}\boldsymbol\beta$$

而

$$\frac{1}{\lambda}(\boldsymbol I_q-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})\boldsymbol U_{\boldsymbol V}\boldsymbol\beta=\frac{1}{\lambda}\left[\boldsymbol\beta^{\mathrm T}\otimes(\boldsymbol I-\boldsymbol\beta\boldsymbol\beta^{\mathrm T})\right]\mathrm{vec}(\boldsymbol U_{\boldsymbol V})$$

由定理 4.3.1 和定理 4.3.2，上式的极限分布为多元正态分布，期望为 **0**，协方差阵为式（4.12）。

证毕。

对于式（4.9），只要用 $E(\boldsymbol Z|Y)$ 的矩估计代替式（4.9）中的 $E(\boldsymbol Z|Y)$，便可以得到渐近协方差阵式（4.9）的相合估计。同样，用 $\boldsymbol m_h,\boldsymbol\mu,\boldsymbol\Sigma,\boldsymbol\nu_h,\boldsymbol G,\tilde{\boldsymbol G}_h,\boldsymbol H$ 的矩估计代替 $\boldsymbol m_h,\boldsymbol\mu,\boldsymbol\Sigma,\boldsymbol\nu_h,\boldsymbol G,\tilde{\boldsymbol G}_h,\boldsymbol H$ 也可以得到式（4.10）的相合估计。然后，用 $\lambda,\boldsymbol\beta$ 的估计 $\hat\lambda,\hat{\boldsymbol\beta}$ 代替式（4.12）中的 $\lambda,\boldsymbol\beta$，就可以得到回归系数 $\boldsymbol\beta$ 的渐近协方差阵的相合估计。

4.4 模拟研究

4.4.1 点估计

模拟一：

首先，对由线性模型 $\boldsymbol{Y}^* = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 和映射 $\boldsymbol{Y} = \arg\max\limits_{h}\{y_1^*, y_2^*, \cdots, y_p^*\}$ 确定的多项选择模型进行模拟。

取 $\boldsymbol{X}$ 为 4×5 的随机矩阵，$\mathrm{vec}(\boldsymbol{X}) \sim N_{20}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_x)$，

$$\boldsymbol{\mu} = 0.1\times(1, 2, 3, 4; 4, 3, 2, 1; 6, 1, 2, 1; 2, 0, 7, 1; 0, 5, 2, 2)^{\mathrm{T}},$$

回归系数真值取 $\boldsymbol{\beta} = (1, 2, 0, 1)^{\mathrm{T}}$。$\boldsymbol{\Sigma}_x$ 对角线上元素为 1，其余元素均为 0.2。扰动项 $\boldsymbol{\varepsilon} \sim N_4(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$，协方差阵 $\boldsymbol{\Sigma}_\varepsilon$ 对角线上元素为 1，其余元素均为 0.2。分别产生每组样本容量为 50，100，500 的随机数进行模拟。实验重复 500 次，将估计的均值列入表 4.1 中。使用两个评价标准：一是结果的每个分量与真值的均方误差；二是结果与真值夹角的余弦，即

$$\cos(\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}) = \frac{\langle\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}\rangle}{\langle\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}\rangle^{\frac{1}{2}}\langle\boldsymbol{\beta}, \boldsymbol{\beta}\rangle^{\frac{1}{2}}}$$

模拟结果如表 4.1 和表 4.2 所示。括号内为分量的均方误差和夹角余弦值的标准差。

表 4.1　估计的均值及均方误差（一）

n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\cos(\hat{\beta}, \beta)$
50	0.4004	0.8247	−0.0241	0.3986	0.9559
	（0.1054）	（0.0893）	（0.1178）	（0.1245）	（0.0487）
100	0.4027	0.8135	−0.0072	0.3994	0.9909
	（0.0669）	（0.0304）	（0.0740）	（0.0674）	（0.0076）
500	0.4086	0.8162	−0.0014	0.4084	0.9986
	（0.0249）	（0.0111）	（0.0308）	（0.0263）	（0.0011）

表 4.2　对应样本矩阵的特征值（一）

n=50	2.5511	0.9335	0.5532	0.3105
n=100	2.0493	0.3098	0.1898	0.1076
n=500	1.8485	0.0478	0.0296	0.0174

逆回归方法是一种大样本方法，模拟结果表明，在大样本情况下，估计很准确，即使在样本容量较小的情况下（n=50），也得到了不错的估计结果，这

反映了逆回归方法对于由线性回归方程确定的多项 Probit 模型是一种比较有效的估计方法。

表 4.2 中最大特征值远远大于其他特征值，而第二特征值越小，说明最大特征值对应的特征向量越显著，估计的误差越小。

模拟二：

第二个模拟中模型为式（4.16）的回归方程和映射式（1.3）所确定的 GMNC 模型。

$$y_h^* = \boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta}(1+0.5\boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta})+\varepsilon_h \tag{4.16}$$

在模拟过程中，自变量与扰动项的分布与模拟一相同，回归系数真值仍取 $\boldsymbol{\beta}=(1,2,0,1)$。为了更好地说明估计的性质，这里样本容量增加为 1000 的模拟估计。模拟结果如表 4.3 和表 4.4 所示，括号内为分量的均方误差和夹角余弦值的标准差。

表 4.3　估计的均值与均方误差（二）

n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\cos(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})$
50	0.30242	0.7434	−0.1271	0.2146	0.5507
	（0.3752）	（0.3163）	（0.3596）	（0.3128）	（0.4309）
100	0.3630	0.7920	−0.0616	0.3101	0.8035
	（0.1831）	（0.2350）	（0.3301）	（0.3831）	（0.1801）
500	0.4134	0.8170	−0.0122	0.4086	0.9871
	（0.0839）	（0.0486）	（0.0191）	（0.0981）	（0.0128）
1000	0.4081	0.8181	−0.0433	0.4054	0.9976
	（0.0410）	（0.0273）	（0.0472）	（0.0566）	（0.0021）

表 4.4　对应样本矩阵的特征值（二）

n=50	1.5597	1.0304	0.5837	0.3863
n=100	0.6824	0.4019	0.2641	0.1546
n=500	0.3218	0.0687	0.0433	0.0267
n=1000	0.2983	0.0313	0.0198	0.0115

模拟结果分析如下。

从模拟的结果来看，小样本时估计不是很准确，当 n=50 时，无论是从各分量的估计来看，还是从夹角余弦还看，估计误差都比较大。例如，夹角余弦的均值为 $\cos(\hat{\boldsymbol{\beta}},\boldsymbol{\beta})=0.5507$，当样本量较小时，估计的方向误差较大，估计的方向与参数真值方向的夹角的平均误差超过了 25°。因为这个模型不是线性模型，而是一个比较复杂的非线性模型，所以样本容量较小时估计不够准确。

随着样本容量的增大，估计渐渐向真值方向收敛，从 n=100 到 n=500、100，

模拟结果存在一个明显的收敛趋势，说明估计是渐近正态的。

对于 n=1000 的大样本，模拟结果显示出的估计还是很好的，夹角余弦的均值为 $\cos(\hat{\beta}, \beta) = 0.9976$，说明得到的估计方向与真值方向非常接近，表明逆回归方法在回归模型为非线性模型时也可以得到很好的估计。

对于非线性模型而言，当因变量与自变量存在一个明确的函数关系，而且已知回归方程的具体形式和扰动项的具体分布时，当样本容量较小时可以用最小二乘法拟合回归方程。最小二乘法的估计效率比较高，但其使用条件比较苛刻，在很多时候很难满足使用条件。而多项选择模型中根本不存在自变量与因变量的直接函数关系，因此根本无法直接应用最小二乘法求回归系数。而当数据量较大时，可以应用本节所述方法求回归系数方法的估计，获得模型的信息。

4.4.2　假设检验

1．检验犯第一类错误频率的模拟

这里取模拟二中的数据进行线性假设检验。检验的假设为回归系数的线性假设

$$H_0\text{:}\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta} = 0$$

$$H_1\text{:}\ \boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta} \neq 0$$

式中

$$\boldsymbol{L}^{\mathrm{T}} = \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & -1 & 0 & 2 \end{bmatrix}$$

令检验统计量为

$$\boldsymbol{T} = n\hat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{L}(\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\Omega}}_{\gamma}\boldsymbol{L})^{-1}\boldsymbol{L}^{\mathrm{T}}\hat{\boldsymbol{\beta}}$$

式中，$\hat{\boldsymbol{\Omega}}_{\gamma}$ 为渐近协方差阵的估计。

这里我们给出自变量 $\boldsymbol{X}$ 的协方差阵未知时检验的水平和势的模拟结果。表 4.5 为检验犯第一类错误的频率，样本容量取 $n = 500, 1000, 1500$。

表 4.5　检验犯第一类错误的频率

n	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.025$
500	0.1363	0.0815	0.0454
1000	0.1118	0.0577	0.0367
1500	0.0964	0.0488	0.0282

2．模拟结果分析

这里假设检验的模拟主要考察本章给出回归系数估计的渐近理论是否正

确。基于前面点估计的模拟结果，已知对于式（4.16）这样复杂的回归函数，小样本估计不够精确，所以这里对于假设检验的模拟直接选择了比较大的样本容量。样本容量分别取 n=500，1000，1500。

当 n=500 时，模拟结果表明检验犯第一类错误的概率大于给定的显著性水平，说明在估计渐近协方差 $\hat{\boldsymbol{\Omega}}_\gamma$ 时需要的样本容量更大，只有 500 个数据时，不足以得到渐近协方差阵 $\hat{\boldsymbol{\Omega}}_\gamma$ 比较准确的估计。因此造成犯第一类错误的频率比较大。

当我们放大样本容量时，即当 n=1000 时，发现检验犯第一类错误的频率已经降低，开始向真实水平接近，但仍然要比真实水平大一些。

当样本容量为 n=1500 时，检验犯第一类错误的频率已经接近检验的真实水平，说明这时的渐近协方阵的估计是比较准确的。

从模拟结果能够观察出来，回归系数估计的渐近正态性是存在的，并且给出的渐近协方差阵也是正确的，只不过当模型比较复杂时，会影响到估计与其渐近协方差阵的收敛速度。

3．检验的势模拟

下面模拟考察上述检验的势。回归系数 $\boldsymbol{\beta}$ 的真值分别取以下不同的值。

$\boldsymbol{\beta}=(1, 2.25, 0, 1)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1, 2, 0, 1.25)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1, 2.5, 0, 1)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1, 2, 0, 1.5)^{\mathrm{T}}$, $\boldsymbol{\beta}=(1, 1, 0, 1)^{\mathrm{T}}$ $\boldsymbol{L}$ 的取值仍为

$$\boldsymbol{L}^{\mathrm{T}}=\begin{bmatrix}1 & -1 & 1 & 1\\ 0 & -1 & 0 & 2\end{bmatrix}$$

满足

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=\begin{bmatrix}-0.25\\ -0.25\end{bmatrix},\begin{bmatrix}0.25\\ 0.5\end{bmatrix},\begin{bmatrix}-0.5\\ -0.5\end{bmatrix},\begin{bmatrix}0.5\\ 1\end{bmatrix},\begin{bmatrix}1\\ 1\end{bmatrix}$$

模拟结果如表 4.6 所示。

表 4.6　检验的势，显著性水平 $\alpha=0.05$

$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	$(-0.25,-0.25)^{\mathrm{T}}$	$(0.25,0.5)^{\mathrm{T}}$	$(-0.5,-0.5)^{\mathrm{T}}$	$(0.5,1)^{\mathrm{T}}$	$(1,1)^{\mathrm{T}}$
n=500	0.2893	0.3288	0.6120	0.6379	0.8463
n=1000	0.3687	0.6752	0.8373	0.8967	0.9710
n=1500	0.5258	0.8936	0.9374	1	1

4．模拟结果的分析

模拟仍然是在大样本情形 n=500，1000，1500 下进行的。模拟结果表明，当样本容量增大时，在同一检验下，如取值为 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}=(-0.25,-0.25)^{\mathrm{T}}$ 检验，随着样本容量的增加，势函数的值随之变大，说明增加的样本容量能提供信息使检

验变得更加精确。另外，随着 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$ 取值各分量绝对值的增大，距离（0，0）点越远，检验的势越大，说明估计是渐近正态的。

下面模拟 $\boldsymbol{L}$ 取不同值时，不同检验的势，$\boldsymbol{L}^{\mathrm{T}}$ 的取值为下列矩阵：

$$\begin{bmatrix} 1 & -1 & 1 & 1.5 \\ 0 & -1 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & -1.25 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 1 & 1.5 \\ 0 & -1.25 & 0 & 2 \end{bmatrix},$$

$$\begin{bmatrix} 1 & -1 & 1 & 1 \\ 0 & -1.5 & 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & -1 & 1 & 2 \\ 0 & -1.5 & 0 & 2 \end{bmatrix}$$

满足

$$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta} = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -0.5 \end{bmatrix}, \begin{bmatrix} -0.5 \\ -0.5 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

模拟结果如表 4.7 所示。

表 4.7　检验的势（$\alpha = 0.05$）

$\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$	$(0.5,0)^{\mathrm{T}}$	$(0,-0.5)^{\mathrm{T}}$	$(0.5,-0.5)^{\mathrm{T}}$	$(0,-1)^{\mathrm{T}}$	$(1,-1)^{\mathrm{T}}$
n=500	0.4475	0.3712	0.7518	0.9311	1
n=1000	0.8857	0.6795	0.9786	1	1
n=1500	0.9845	0.9266	1	1	1

模拟仍然是在大样本情形 n=500，1000，1500 下进行的。当 $\boldsymbol{L}$ 取不同值时，实际是进行不同的检验。模拟结果也显示出了这一点，当样本容量增大时，在同一检验下，如取值为 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta} = (0.5, 0)^{\mathrm{T}}$ 检验，随着样本容量的增加，势函数的值随之变大，同样说明增加的样本容量能提供信息使检验变得更加准确。不同的检验，即使 $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{\beta}$ 的取值关于（0，0）点是距离相等的，但检验的势也不同。

第 5 章　多重二元响应 Probit 模型的渐近有效估计

5.1　多重二元响应 Probit 模型

首先，回顾一下多重二元响应模型。由线性模型 $\boldsymbol{Y}^* = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}(\boldsymbol{X} \in \mathbf{R}^{q\times p}, \boldsymbol{\beta} \in \mathbf{R}^q, \boldsymbol{\varepsilon} \in \mathbf{R}^p)$ 和映射 $\boldsymbol{Y} = \tau(\boldsymbol{Y}^*)$ 确定的 LDV 模型称为多重二元响应模型，其中映射 τ 满足

$$y_j = \begin{cases} 1, & y_j^* > 0 \\ 0, & y_j^* \leqslant 0, \ j = 1, \cdots, p \end{cases}$$

即 $\boldsymbol{Y} = (y_1, \cdots, y_p)^{\mathrm{T}}$ 是一个由 0 和 1 构成的 p 维列向量。当扰动项 $\boldsymbol{\varepsilon}$ 服从多元正态分布，即 $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\varOmega})$ 时，称上述模型为多重二元响应 Probit 模型。

如果 $\boldsymbol{\varepsilon}$ 的各分量独立，$\varepsilon_i \sim N(0, \sigma_i^2)\ (i = 1, 2, \cdots, p)$，这时模型为二元响应 Probit 模型。本章研究的多重二元响应模型 $\boldsymbol{\varepsilon}$ 的各分量可以是不独立的。

针对多重二元响应 Probit 模型，提出了二步估计方法。第一步由边际似然得到参数 $\sqrt{n}$ 相合的估计，第二步通过一步迭代（或较少的几步迭代）得到渐近有效估计。5.2 节给出了参数的边际似然估计；5.3 节给出了多重二元响应模型的 Fisher 信息阵；5.4 节给出了参数的渐近有效估计及相关的理论证明；5.5 节给出了模拟结果。

5.2　边际似然估计

对于参数有可识别限制的多重二元响应 Probit 模型，记 $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\varOmega})$，其中

$$\boldsymbol{\varOmega} = \begin{bmatrix} 1 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ & \cdots & \cdots & \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{bmatrix}$$

待估参数记为

$$\boldsymbol{\theta}=(\beta_1,\cdots,\beta_q,\sigma_2^2,\cdots,\sigma_p^2,\sigma_{12},\cdots,\sigma_{1p},\cdots,\sigma_{p-1p})^{\mathrm{T}}$$

参数 $\boldsymbol{\theta}$ 是一个 $q+\dfrac{p\times(p+1)}{2}-1$ 维的向量。

$\boldsymbol{y}_1^*$ 的边际分布为 $N(\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta},1)$，$\boldsymbol{y}_j^*$ 的边际分布为 $N(\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta},\boldsymbol{\sigma}_j^2)(j=2,\cdots,p)$，$(\boldsymbol{y}_h^*,\boldsymbol{y}_j^*)$ 的联合边际分布为二元正态分布，即

$$N\left(\begin{bmatrix}\boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta}\\ \boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}\end{bmatrix},\begin{bmatrix}\boldsymbol{\sigma}_h^2 & \boldsymbol{\sigma}_{hj}\\ \boldsymbol{\sigma}_{jh} & \boldsymbol{\sigma}_j^2\end{bmatrix}\right)$$

其中，$\boldsymbol{X}_j$ 表示矩阵 $\boldsymbol{X}$ 的第 j 列元素，$\sigma_1^2=1$；$h\neq j$；h、$j=1,2,\cdots,p$。

令 $\boldsymbol{D}(y_j)$ 表示给定 y_j 时潜在变量 y_j^* 的逆像：

$$\boldsymbol{D}(y_j)=\{y_j^*:y_j=\tau(y_j^*)\}=\begin{cases}y_j^*\in(0,+\infty), y_j=1,\\ y_j^*\in(-\infty,0],\quad y_j=0,\end{cases}\quad j=1,2,\cdots,p$$

$\boldsymbol{D}(y_h,y_j)$ 表示给定 (y_h,y_j) 时潜在变量 (y_h^*,y_j^*) 的逆像：

$$\begin{aligned}\boldsymbol{D}(y_h,y_j)&=\{(y_h^*,y_j^*):(y_h,y_j)=\tau(y_h^*,y_j^*)\}\\&=\begin{cases}y_h^*\in(0,+\infty), y_j^*\in(0,+\infty), & (y_h,y_j)=(1,1)\\ y_h^*\in(0,+\infty), y_j^*\in(-\infty,0], & (y_h,y_j)=(1,0)\\ y_h^*\in(-\infty,0], y_j^*\in(0,+\infty), & (y_h,y_j)=(0,1)\\ y_h^*\in(-\infty,0], y_j^*\in(-\infty,0], & (y_h,y_j)=(0,0)\end{cases}\end{aligned}$$

$$h\neq j;h\text{、}j=1,\cdots,p$$

则 y_1,y_j 和 (y_h,h_j) 的边际概率分布分别为

$$P_1(y_1;\boldsymbol{X}_1,\boldsymbol{\beta})=\int_{y_1^*\in D(y_1)}n_1(y_1^*-\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta},1)\mathrm{d}y_1^* \tag{5.1}$$

$$P_j(y_j;\boldsymbol{X}_j,\boldsymbol{\beta},\sigma_j^2)=\int_{y_j^*\in D(y_j)}n_1(y_j^*-\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta},\sigma_j^2)\mathrm{d}y_j^*,\ j=2,\cdots,p \tag{5.2}$$

$$\begin{aligned}&P_{hj}(y_h,y_j;\boldsymbol{X}_h,\boldsymbol{X}_j,\boldsymbol{\beta},\sigma_h^2,\sigma_j^2,\sigma_{hj})\\&=\int_{(y_h^*,y_j^*)\in\boldsymbol{D}(y_h,y_j)}n_2\left(\begin{bmatrix}y_h-\boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta}\\ y_j-\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)\mathrm{d}y_h^*\mathrm{d}y_j^*,\ h\text{、}j=1,2,\cdots,p\end{aligned} \tag{5.3}$$

其中，$n_T(\cdot,\cdot)$ 是 T 维正态分布的密度函数，如下：

$$n_T(\boldsymbol{z},\boldsymbol{\Omega})=(2\pi)^{-T/2}|\boldsymbol{\Omega}|^{-1/2}\exp[-\boldsymbol{z}^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{z}/2]$$

式中，$\boldsymbol{\Omega}$ 为协方差阵。

边际概率分布 $P_1(y_1;\boldsymbol{X}_1,\boldsymbol{\beta})$ 关于参数 $\boldsymbol{\beta}$ 求导可得

$$\frac{\partial P_1(y_1;\boldsymbol{X}_1,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=P_1(y_1;\boldsymbol{X}_1,\boldsymbol{\beta})E\left\{(y_1^*-\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta})|\ y_1^*\in\boldsymbol{D}(y_1)\right\}\boldsymbol{X}_1 \tag{5.4}$$

由式（5.4）可以得到边际概率分布的得分函数为

$$\begin{aligned}&s_{1\beta}(\boldsymbol{\beta};y_1,\boldsymbol{X}_1)\\&=\frac{\partial\ln P_1(y_1;\boldsymbol{X}_1,\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\\&=E\left\{(y_1^*-\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta})|\ y_1^*\in\boldsymbol{D}(y_1)\right\}\boldsymbol{X}_1\end{aligned} \tag{5.5}$$

如果存在独立观测 $(\boldsymbol{Y}^{(i)},\boldsymbol{X}^{(i)})(i=1,\cdots,n)$，求下式的最大解可以得到 $\boldsymbol{\beta}$ 的边际似然估计。

$$\tilde{\boldsymbol{\beta}}=\arg\max_{\beta}\prod P_1(y_1^{(i)};\boldsymbol{X}_1^{(i)},\boldsymbol{\beta})$$

或者，等价地，求解式（5.6）边际分布的对数似然方程也可以得到 $\boldsymbol{\beta}$ 的边际似然估计。

$$\sum_i s_{1\beta}(\boldsymbol{\beta};y_1^{(i)},\boldsymbol{X}_1^{(i)})=0 \tag{5.6}$$

其中

$$s_{1\beta}(\boldsymbol{\beta};y_1^{(i)},\boldsymbol{X}_1^{(i)})=\frac{\int_{y_1^{*(i)}\in\boldsymbol{D}}(y_1^{*(i)}-\boldsymbol{X}_1^{(i)\mathrm{T}}\boldsymbol{\beta})n_1(y_1^{*(i)}-\boldsymbol{X}_1^{(i)\mathrm{T}}\boldsymbol{\beta},1)\mathrm{d}y_1^{*(i)}}{\int_{y_1^{*(i)}\in\boldsymbol{D}}n_1(y_1^{*(i)}-\boldsymbol{X}_1^{(i)\mathrm{T}}\boldsymbol{\beta},1)\mathrm{d}y_1^{*(i)}}\boldsymbol{X}_1^{(i)}$$

边际概率分布 $P_{hj}(y_h,y_j;\boldsymbol{X}_h,\boldsymbol{X}_j,\boldsymbol{\beta},\sigma_h^2,\sigma_j^2,\sigma_{hj})$ 关于 σ_h^2 的偏导数为

$$\begin{aligned}&\frac{\partial P_{hj}(y_h,y_j;\boldsymbol{X}_h,\boldsymbol{X}_j,\sigma_{hj},\sigma_h^2,\sigma_j^2,\boldsymbol{\beta})}{\partial\sigma_h^2}\\=\ &-\frac{1}{2}P_{hj}(y_h,y_j;\sigma_{hj},\sigma_h^2,\sigma_j^2,\boldsymbol{\beta})\sigma_h^2\nu\Big\{1-E\Big[\nu(u_h^2\sigma_j^2+u_2^2\sigma_j^2-\\&2u_hu_j\sigma_{hj})-u_j^2\sigma_h^{-2}|(y_h^*,y_j^*)\in\boldsymbol{D}(y_h,y_j)\Big]\Big\}\end{aligned}$$

式中，$\nu=(\sigma_h^2\sigma_j^2-\sigma_{hj}^2)^{-1}$，$u_h=y_h^*-\boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta},u_j=y_j^*-\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}$。

σ_h^2 关于边际分布的得分函数记为 $s_{hj\sigma_{hj}}(\sigma_{hj},\sigma_h^2,\sigma_j^2,\boldsymbol{\beta};y_h,y_j,\boldsymbol{X}_h,\boldsymbol{X}_j)$，可得

$$\begin{aligned}&s_{hj\sigma_h^2}(\sigma_{hj},\sigma_h^2,\sigma_j^2,\boldsymbol{\beta};y_h,y_j,\boldsymbol{X}_h,\boldsymbol{X}_j)\\&=-\frac{1}{2}\sigma_h^2\nu\Big\{1-E\Big[\nu(u_h^2\sigma_j^2+u_j^2\sigma_h^2-2u_hu_j\sigma_{hj})-u_j^2\sigma_h^{-2}|(y_h^*,y_j^*)\in\boldsymbol{D}(y_h,y_j)\Big]\Big\}\end{aligned}$$

则 σ_h^2 的边际似然方程为

$$\sum_i s_{hj\sigma_h^2}(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)}) = 0 \tag{5.7}$$

其中

$$s_{hj\sigma_h^2}(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)})$$

$$= -\frac{1}{2}\sigma_h^2\nu\left\{1-\frac{\int_{(y_h^{*(i)},\ y_j^{*(i)})\in\boldsymbol{D}(y_h^{(i)},\ y_j^{(i)})} G_h(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)})\mathrm{d}y_h^*\mathrm{d}y_j^*}{\int_{(y_h^{*(i)},\ y_j^{*(i)})\in\boldsymbol{D}(y_h^{(i)},\ y_j^{(i)})} n_2\left(\begin{bmatrix}u_h^{(i)}\\ u_j^{(i)}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)\mathrm{d}y_h^*\mathrm{d}y_j^*}\right\}$$

$$G_h(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)})$$

$$= \left[\nu(u_h^{(i)^2}\sigma_j^2 + u_j^{(i)^2}\sigma_j^2 - 2u_h^{(i)}u_j^{(i)}\sigma_{hj}) - u_j^{(i)^2}\sigma_h^{-2}\right] n_2\left(\begin{bmatrix}u_h^{(i)}\\ u_j^{(i)}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)$$

而 $u_h^{(i)} = y_h^{*(i)} - \boldsymbol{X}_h^{(i)}\boldsymbol{\beta}, u_j^{(i)} = y_j^{*(i)} - \boldsymbol{X}_j^{(i)\mathrm{T}}\boldsymbol{\beta}$ 。

同理，σ_j^2 的边际似然方程为

$$\sum_i s_{hj\sigma_j^2}(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)}) = 0 \tag{5.8}$$

其中

$$s_{hj\sigma_j^2}(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)})$$

$$= -\frac{1}{2}\sigma_j^2\nu\left\{1-\frac{\int_{(y_h^{*(i)},\ y_j^{*(i)})\in\boldsymbol{D}(y_h^{(i)},\ y_j^{(i)})} G_j(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)})\mathrm{d}y_h^*\mathrm{d}y_j^*}{\int_{(y_h^{*(i)},\ y_j^{*(i)})\in\boldsymbol{D}(y_h^{(i)},\ y_j^{(i)})} n_2\left(\begin{bmatrix}u_h^{(i)}\\ u_j^{(i)}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)\mathrm{d}y_h^*\mathrm{d}y_j^*}\right\}$$

$$G_j(\sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta};\ y_h^{(i)},\ y_j^{(i)},\ \boldsymbol{X}_h^{(i)},\ \boldsymbol{X}_j^{(i)})$$

$$= \left[\nu(u_h^{(i)^2}\sigma_j^2 + u_j^{(i)^2}\sigma_j^2 - 2u_h^{(i)}u_j^{(i)}\sigma_{hj}) - u_h^{(i)^2}\sigma_j^{-2}\right] n_2\left(\begin{bmatrix}u_h^{(i)}\\ u_j^{(i)}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)$$

概率分布式（5.3）关于 σ_{hj} 的偏导数为

$$\frac{\partial P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j,\ \sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta})}{\partial\sigma_{hj}}$$

$$= \quad P_{hj}(y_h, y_j;\ \sigma_{hj},\ \sigma_h^2,\ \sigma_j^2,\ \boldsymbol{\beta})\sigma_{hj}\nu$$

$$\left\{1 - E\left[\nu(u_h^2\sigma_j^2 + u_2^2\sigma_j^2 - 2u_h u_j\sigma_{hj}) - u_h u_j\sigma_{hj}^{-1}\middle|(y_h^*, y_j^*)\in\boldsymbol{D}(y_h, y_j)\right]\right\}$$

式中，$\nu=(\sigma_h^2\sigma_j^2-\sigma_{hj}^2)^{-1}$，$u_h=y_h^*-\boldsymbol{X}_h^{\mathrm{T}}\boldsymbol{\beta}, u_j=y_j^*-\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}$。

σ_{hj} 关于边际分布的得分函数记为 $s_{hj\sigma_{hj}}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$，可得

$$\begin{aligned}&s_{hj\sigma_{hj}}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)\\&=\sigma_{hj}\nu\left\{1-E\left[\nu(u_h^2\sigma_j^2+u_j^2\sigma_h^2-2u_hu_j\sigma_{hj})-u_hu_j\sigma_{hj}^{-1}\middle|(y_h^*, y_j^*)\in\boldsymbol{D}(y_h, y_j)\right]\right\}\end{aligned}$$

则 σ_{hj} 的边际似然方程为

$$\sum_i s_{hj\sigma_{hj}}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})=0 \tag{5.9}$$

其中

$$\begin{aligned}&s_{hj\sigma_{hj}}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})\\&=\sigma_{hj}\nu\left\{1-\frac{\int_{(y_h^{*(i)},\ y_j^{*(i)})\in\boldsymbol{D}(y_h^{(i)},\ y_j^{(i)})}G_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})\mathrm{d}y_h^*\mathrm{d}y_j^*}{\int_{(y_h^{*(i)},\ y_j^{*(i)})\in\boldsymbol{D}(y_h^{(i)},\ y_j^{(i)})}n_2\left(\begin{bmatrix}u_h^{(i)}\\u_j^{(i)}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)\mathrm{d}y_h^*\mathrm{d}y_j^*}\right\}\end{aligned}$$

$$\begin{aligned}&G_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})\\&=\left[\nu(u_h^{(i)}2\sigma_j^2+u_j^{(i)^2}\sigma_j^2-2u_h^{(i)}u_j^{(i)}\sigma_{hj})-u_h^{(i)}u_j^{(i)}\sigma_{hj}^{-1}\right]n_2\left(\begin{bmatrix}u_h^{(i)}\\u_j^{(i)}\end{bmatrix},\begin{bmatrix}\sigma_h^2 & \sigma_{hj}\\ \sigma_{jh} & \sigma_j^2\end{bmatrix}\right)\end{aligned}$$

本节我们假定参数的边际极大似然估计存在，可以按照下面的步骤得到 $\boldsymbol{\theta}$ 的一个估计：

（1）由式（5.6）得到 $\boldsymbol{\beta}$ 的边际似然估计 $\tilde{\boldsymbol{\beta}}$。

（2）将 $\tilde{\boldsymbol{\beta}}$ 代入由式（5.7）～式（5.9）组成的方程组中，得到 σ_j^2 和 σ_{hj} 的边际似然估计 $\tilde{\sigma}_j^2\ (j=2,\cdots,p)$ 和 $\tilde{\sigma}_{hj}\ (h\neq j; h=1, 2,\cdots,p;\ j=2,\cdots,p)$。

注 5.2.1　当 $\boldsymbol{\beta}$ 未知时，由概率分布式（5.3）确定的统计模型中的参数仍是不可识别的，这里第一步得到 $\boldsymbol{\beta}$ 的边际估计；而当 $\boldsymbol{\beta}$ 已知时，余下的参数是可以识别的。

注 5.2.2　利用式（5.7）～式（5.9）组成的方程组求解 $\tilde{\sigma}_j^2\ (j=2,\cdots,p)$ 和 $\tilde{\sigma}_{hj}\ (h\neq j; h=1, 2,\cdots,p;\ j=2,\cdots,p)$ 时，对于同一个 σ_j^2 可以得到 $p-1$ 个解，我们可以用这 $p-1$ 个解的平均值作为 σ_j^2 的边际估计 $\tilde{\sigma}_j^2$。

注 5.2.3　这里的边际似然估计只涉及一维和二维积分，而不涉及三维及以上的高维积分，因此在计算上可以用一般的积分方法解决，而不必利用模拟方

法计算积分。在许多应用软件中都提供了一维、二维积分函数程序，这些程序计算速度比较快。

记 $\tilde{\boldsymbol{\theta}}=(\tilde{\boldsymbol{\beta}}^{\mathrm{T}},\ \tilde{\sigma}_2,\cdots,\ \tilde{\sigma}_p,\ \tilde{\sigma}_{12},\cdots,\ \tilde{\sigma}_{p-1,\,p})^{\mathrm{T}}$，称为 $\boldsymbol{\theta}$ 的边际似然估计（Marginal Maximum Likelihood Estimate，MMLE）。5.4 节中将证明，在满足一定的正则条件下，MMLE 存在并且是 $\sqrt{n}$ 相合的。

5.3 Fisher 信息阵

参数 $\boldsymbol{\theta}$ 的 Fisher 信息阵 $\boldsymbol{I}(\boldsymbol{\theta})$ 中的元素为

$$\boldsymbol{I}_{ij}(\boldsymbol{\theta})=E[s_{\theta_i}(\boldsymbol{Y},\boldsymbol{X})\cdot s_{\theta_j}(\boldsymbol{Y},\boldsymbol{X})],\ i、j=1,\cdots,q+(p-1)+\frac{p(p-1)}{2} \quad (5.10)$$

式中，$s_{\theta_i}(\boldsymbol{Y},\boldsymbol{X})$ 为得分函数 $\boldsymbol{S}(\boldsymbol{\theta})$ 中的分量。

$$\begin{aligned}\boldsymbol{s}_{\boldsymbol{\beta}}(\boldsymbol{Y},\boldsymbol{X}) &= \frac{\partial \ln \boldsymbol{P}(\boldsymbol{Y};\boldsymbol{X},\ \boldsymbol{\theta})}{\partial \boldsymbol{\beta}}\\ &= E\Big[\boldsymbol{X}\boldsymbol{\Omega}^{-1}(\boldsymbol{Y}^*-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})\big|\ \boldsymbol{D}(\boldsymbol{Y})\Big]\end{aligned} \quad (5.11)$$

其中

$$\boldsymbol{P}(\boldsymbol{Y};\boldsymbol{X},\ \boldsymbol{\theta})=\int_{\boldsymbol{D}[\boldsymbol{Y}=\tau(\boldsymbol{Y}^*)]} n_p(\boldsymbol{Y}^*-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta},\ \boldsymbol{\Omega})\mathrm{d}\boldsymbol{Y}^* \quad (5.12)$$

$$\begin{aligned}s_{\sigma_j^2}(\boldsymbol{Y},\boldsymbol{X}) &= \frac{\partial \ln P(\boldsymbol{Y};\boldsymbol{X},\ \boldsymbol{\theta})}{\partial \sigma_j^2}\\ &= -\frac{1}{2}E\Big[\mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{E}_j)-(\boldsymbol{Y}^*-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{E}_j\boldsymbol{\Omega}^{-1}(\boldsymbol{Y}^*-\\ &\quad \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})\big|\ \boldsymbol{D}(\boldsymbol{Y})\Big],\ j=2,\cdots,p\end{aligned} \quad (5.13)$$

式中，tr 表示矩阵的迹，$\boldsymbol{E}_j$ 为 p 阶方阵，除了对角线上第 j 个位置为 1 外，其余元素均为 0。

$$\begin{aligned}s_{\sigma_{ij}}(\boldsymbol{Y},\boldsymbol{X}) &= \frac{\partial \ln P(\boldsymbol{Y};\boldsymbol{X},\ \boldsymbol{\theta})}{\partial \sigma_{ij}}\\ &= -\frac{1}{2}E\Big[\mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{E}_{ij})-(\boldsymbol{Y}^*-\\ &\quad \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})^{\mathrm{T}}\boldsymbol{\Omega}^{-1}\boldsymbol{E}_{ij}\boldsymbol{\Omega}^{-1}(\boldsymbol{Y}^*-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta})\big|\ \boldsymbol{D}(\boldsymbol{Y})\Big]\end{aligned} \quad (5.14)$$

式中，$i\neq j(i,j=1,\cdots,p),\ \boldsymbol{E}_{ij}$ 为一 p 阶对称方阵，第 i 行第 j 列与第 j 行第 i

列元素为 1，其余元素为 0。

若有一组独立观测 $(\boldsymbol{Y}^{(k)}, \boldsymbol{X}^{(k)})(k=1,\cdots,n)$，则 $\boldsymbol{S}(\boldsymbol{\theta})$ 和信息阵中元素 $\boldsymbol{I}_{ij}(\boldsymbol{\theta})$ 可以由式（5.15）和式（5.16）计算

$$\hat{\boldsymbol{S}}(\boldsymbol{\theta})=\frac{1}{n}\sum_{k=1}^{n}\frac{\partial \ln p(\boldsymbol{Y}^{(i)};\boldsymbol{X}^{(i)},\ \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{5.15}$$

$$\hat{\boldsymbol{I}}_{ij}(\boldsymbol{\theta})=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{\partial \ln P(\boldsymbol{Y}^{(k)};\boldsymbol{X}^{(k)},\ \boldsymbol{\theta})}{\partial \theta_i}\cdot\frac{\partial \ln P(\boldsymbol{Y}^{(k)};\boldsymbol{X}^{(k)},\ \boldsymbol{\theta})}{\partial \theta_j}\right) \tag{5.16}$$

式中，i、$j=1, 2,\cdots, q+(p-1)+\dfrac{p(p-1)}{2}$，式中偏导的表达式由式（5.11）、式（5.13）和式（5.14）给出。当 $p\geqslant 4$ 时，式中的期望涉及高维积分，这些高维积分利用数值方法计算不准确，一般利用 Monte-Carlo 方法进行模拟积分。本节利用 Gibbs 抽样方法计算式（5.11）、式（5.13）和式（5.14）中的条件期望，式中的参数 $\boldsymbol{\theta}$ 用边际似然估计 $\tilde{\boldsymbol{\theta}}$ 代替，得到 $\hat{\boldsymbol{S}}(\tilde{\boldsymbol{\theta}})$ 和 $\hat{\boldsymbol{I}}(\tilde{\boldsymbol{\theta}})$。关于 LDV 模型的 Gibbs 抽样方法详见文献[17]。

5.4　渐近有效估计

5.2 节中给出的 MMLE 不是渐近有效的，但可以证明是 $\sqrt{n}$ 相合的。假定 $\{(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)}), i=1, 2,\cdots, n\}$ 为 i.i.d.观测，并且回归系数 $\boldsymbol{\beta}$ 的边际概率分布函数式（5.1）满足文献[68]中定理 2.13 和定理 2.14 的正则条件，那么 $\tilde{\boldsymbol{\beta}}$ 是 $\sqrt{n}$ 相合的。下面证明在满足一定的正则条件时，$\tilde{\sigma}_j^2\ (j=2,\cdots,p)$ 和 $\tilde{\sigma}_{hj}\ (h\neq j; h、j=1,\ 2,\cdots, p)$ 是 $\sqrt{n}$ 相合的。

定义 5.4.1　随机等度连续：若 $\zeta_N(\cdot)$ 是关于参数 $\boldsymbol{\theta}$ 的函数，对于每一个 $\varepsilon>0$ 和 $\lambda>0$，存在 $\delta>0$ 和 N_0 使 $N>N_0$ 时有

$$P\left(\sup_{|\boldsymbol{\theta}-\boldsymbol{\theta}^{\mathrm{T}}|<\delta}|\zeta_N(\boldsymbol{\theta})-\zeta_N(\boldsymbol{\theta}^{\mathrm{T}})|>\varepsilon\right)<\lambda$$

成立。其中，$\boldsymbol{\theta}^{\mathrm{T}}\in\Theta, \boldsymbol{\theta}\in\Theta_1$，则称函数 $\zeta_N(\cdot)$ 在 $\Theta_1\subseteq\Theta$ 上是随机等度连续的。

引理 5.4.1（文献[17]中定理 1）　如果 $\zeta_N(\cdot)$ 在 Θ 上是随机等度连续的，其中 Θ 是一个紧凸集，并且存在某个 $\boldsymbol{\theta}^{\circ}\in\Theta$ 使得 $\zeta_N(\boldsymbol{\theta}^{\circ})$ 是随机有界的，则 $\zeta_N(\cdot)$ 在 Θ 上是一致随机有界的。

引理 5.4.2 在多重二元响应 Probit 模型中有 $0 < P(y_j = 1) < 1$，记

$$l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) = \sum_i \ln P_{hj}(y_h^{(i)}, y_j^{(i)}; \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})$$

$$l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) = \sum_i \ln P_{hj}(y_h^{(i)}, y_j^{(i)}; \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}})$$

式中，$h \neq j, h、j = 1, \cdots, p$，$\tilde{\beta}$ 是由式（5.6）得到的边际似然估计，并且 $\tilde{\beta}$ 是 $\sqrt{n}$ 相合的。如果模型满足下述条件：

（1）记 $\boldsymbol{\beta}$ 的参数空间为 Θ_β, Θ_β 是 $\mathbf{R}^q$ 上的开矩形，并且真值 $\boldsymbol{\beta}^\circ$ 是 Θ_β 的内点。

（2）在参数真值 $\boldsymbol{\beta}^\circ$ 的邻域内，对几乎所有的 $(\boldsymbol{Y}, \boldsymbol{X})$，有 $\ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})$ 关于 $\boldsymbol{\beta}$ 存在各阶偏导数，并且有

$$E_{(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \beta^\circ)}\left(\frac{\partial \ln P_{hj}(y_h, y_j; X_h, X_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_k}\right) < \infty, k = 1, \cdots, q$$

得分函数及其导数可由一个独立于 $\boldsymbol{\theta}$ 的，并且存在有限一、二阶矩的函数所控。

（3）记

$$B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_j^{(i)}, \boldsymbol{X}_j^{(i)}) = \frac{\partial^2 \ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}, k、l = 1, \cdots, q$$

为 Hession 阵，有

$$E_{(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \beta^\circ)}\left(\frac{\partial^2 \ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}\right) < \infty$$

并且在真值 $\boldsymbol{\beta}_0$ 的邻域内有

$$\frac{\partial^2 \ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}, k、l = 1, \cdots, q$$

关于 $\boldsymbol{\beta}$ 是随机等度连续的。则

$$\frac{1}{n} l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) \to^P E_{(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \beta^\circ)} \ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}) \tag{5.17}$$

式中，$\to^P$ 表示依概率收敛。

证明： 因为 $l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$ 在真值 $\boldsymbol{\beta}^\circ$ 的邻域内存在各阶偏导数，将 $l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$ 在 $\boldsymbol{\beta}^\circ$ 处作二阶 Taylor 展开，得

$$
\begin{aligned}
& l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) \\
= \; & l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) + \\
& \frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \boldsymbol{\beta}}\Big|_{\beta=\beta^{\circ}}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ}) + \\
& \frac{1}{2}\sum_{i=1}^{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ})^{\mathrm{T}} B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^*; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ})
\end{aligned}
$$

其中，$\boldsymbol{\beta}^*$ 落在由 $\tilde{\boldsymbol{\beta}}$ 和 $\boldsymbol{\beta}^{\circ}$ 确定的线段上。

$$
\begin{aligned}
& B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^*; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}) \\
= & \frac{\partial^2 \ln P_{hj}(y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}\Big|_{\beta=\beta^*}
\end{aligned}
$$

$$k、l = 1, \cdots, q$$

所以有

$$
\begin{aligned}
& \frac{1}{n} l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) - \frac{1}{n} l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) \\
= \; & \frac{1}{n}\frac{1}{\sqrt{n}} \frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \boldsymbol{\beta}}\Big|_{\beta=\beta^{\circ}} \sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ}) + \\
& \frac{1}{2}\frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ})^{\mathrm{T}} B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^*; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ})
\end{aligned}
$$

由于 $\tilde{\boldsymbol{\beta}}$ 是 $\sqrt{n}$ 相合的，则有

$$\sqrt{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ}) = O_p(1)$$

又由 $\dfrac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \boldsymbol{\beta}}\Big|_{\beta=\beta^{\circ}}$ 为 i.i.d.变量和，所以有

$$\frac{1}{\sqrt{n}} \frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \boldsymbol{\beta}}\Big|_{\beta=\beta^{\circ}} = O_p(1)$$

又因为

$$
\begin{aligned}
& \sum_{i=1}^{n}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ})^{\mathrm{T}} B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^*; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^{\circ}) \\
= & \sum_{i=1}^{n}\sum_{k=1}^{q}\sum_{l=1}^{q} \frac{\partial^2 \ln P_{hj}(y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}\Big|_{\beta=\beta^*}(\tilde{\beta}_k - \beta_k^{\circ})(\tilde{\beta}_l - \beta_l^{\circ})
\end{aligned}
$$

由

$$E_{(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \beta^{\circ})}\left(\frac{\partial^2 \ln P_{hj}(y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \beta_h \partial \beta_k}\right) < \infty$$

和

$$\frac{\partial^2 \ln P_{hj}(y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial\beta_h\partial\beta_k}$$

关于 $\boldsymbol{\beta}$ 的随机等度连续性可得

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln P_{hj}(y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)}; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial\beta_h\partial\beta_k}\bigg|_{\beta=\beta^*}$$

依据概率有界，有

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^\circ)^{\mathrm{T}}B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^*; y_h{}^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^\circ)\\
&=\frac{1}{n}\sqrt{n}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^\circ)^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}B(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^*; y_h^{(i)}, y_j^{(i)}, \boldsymbol{X}_h^{(i)}, \boldsymbol{X}_j^{(i)})\sqrt{n}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}^\circ)\\
&=O_p\left(\frac{1}{n}\right)\end{aligned}$$

因此可得

$$\begin{aligned}&\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)-\\
&\quad\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^\circ, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)=O_p\left(\frac{1}{n}\right)\end{aligned}\tag{5.18}$$

又由强大数定理得

$$\begin{aligned}&\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^\circ, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)\to^{a.s.} E_{(\sigma_{hj},\, \sigma_h^2,\, \sigma_j^2,\, \boldsymbol{\beta}^\circ)}\\
&\quad\ln P_{hj}(y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^\circ)\end{aligned}$$

所以有

$$\begin{aligned}&\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)\to^{P} E_{(\sigma_{hj},\, \sigma_h^2,\, \sigma_j^2,\, \boldsymbol{\beta}^\circ)}\\
&\quad\ln P_{hj}(y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j; \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^\circ)\end{aligned}$$

证毕。

我们注意到，式（5.18）可以得到

$$\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)=\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^\circ, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)+O_p\left(\frac{1}{n}\right)$$

如果 $l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$ 关于 $\boldsymbol{\vartheta}_{hj}=(\sigma_{hj}, \sigma_h^2, \sigma_j^2)^{\mathrm{T}}$ 存在各阶偏导数，并且其各阶偏导数也满足引理 5.4.2 中的条件，同样可以得到下面的结论

$$
\begin{aligned}
&\frac{1}{n}\frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}} \\
&= \frac{1}{n}\frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}} + \\
&O_p\left(\frac{1}{n}\right)\frac{1}{n}\frac{\partial^2 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\partial \vartheta_{hjl}} \\
&= \frac{1}{n}\frac{\partial^2 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\partial \vartheta_{hjl}} + \\
&O_p\left(\frac{1}{n}\right)\frac{1}{n}\frac{\partial^3 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\partial \vartheta_{hjl}\partial \vartheta_{hjm}} \\
&= \frac{1}{n}\frac{\partial^3 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ}, y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\partial \vartheta_{hjl}\partial \vartheta_{hjm}} + O_p\left(\frac{1}{n}\right)
\end{aligned}
\tag{5.19}
$$

式中，ϑ_{hjk}、ϑ_{hjl}、ϑ_{hjm} 表示向量 $\boldsymbol{\vartheta}_{hj}$ 中的分量，k、l、$m=1, 2, 3$。

引理 5.4.3　在多重二元响应 Probit 模型中若有 $0<P(y_j=1)<1$，并且满足以下条件：

（1）$\boldsymbol{\beta}$ 的边际似然估计 $\tilde{\boldsymbol{\beta}}$ 是 $\sqrt{n}$ 相合的。

（2）模型满足引理 5.4.2 中的条件。

（3）令参数

$$
\boldsymbol{\theta}=(\beta_1,\cdots, \beta_q, \sigma_2^2,\cdots, \sigma_p^2, \sigma_{12},\cdots, \sigma_{1p},\cdots, \sigma_{p-1p})^{\mathrm{T}}
$$

的参数空间为 Θ, $\boldsymbol{\vartheta}_{hj}=(\sigma_{hj}, \sigma_h^2, \sigma_j^2)^{\mathrm{T}}(h\neq j; h、j=1, 2,\cdots, p)$，其子参数空间为 Θ_{hj}。存在 Θ_{hj} 的一个开子集 ω_{hj}，它包含真实子参数 $\boldsymbol{\vartheta}_{hj}^{\circ}=(\sigma_{hj}^{\circ}, \sigma_h^{\circ 2}, \sigma_j^{\circ 2})^{\mathrm{T}}$，使几乎所有的 $(\boldsymbol{Y}, \boldsymbol{X})$，$P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})$ $(h\neq j; h、j=1, 2, \cdots, p)$ 对所有 $\boldsymbol{\vartheta}_{hj}\in\omega_{hj}$ 和一个包含参数 $\boldsymbol{\beta}$ 真值的开子集上存在三阶导数

$$
\frac{\partial P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})}{\partial \vartheta_{hjk}\partial \vartheta_{hjl}\partial \vartheta_{hjm}}, k、l、m=1, 2, 3
$$

式中，ϑ_{hjk}、ϑ_{hjl}、ϑ_{hjm} 为参数向量 $\boldsymbol{\vartheta}_{hj}$ 中的分量。

（4）令

$$
E_{\vartheta_{hj}}\left\{\frac{\partial}{\partial \vartheta_{hjk}}\ln\left[P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ})\right]\right\}=0
$$

$$
\begin{aligned}
&\boldsymbol{I}_{kl}(\boldsymbol{\vartheta}_{hj}) \\
&= E_{\vartheta_{hj}}\left\{\frac{\partial}{\partial \vartheta_{hjk}} \ln\left[P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ})\right] \frac{\partial}{\partial \vartheta_{hjl}}\right. \\
&\quad \left.\ln\left[P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ})\right]\right\} \\
&= E_{\vartheta_{hj}}\left\{-\frac{\partial^2}{\partial \vartheta_{hjk} \partial \vartheta_{hjl}} \ln\left[P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ})\right]\right\}
\end{aligned}
$$

$$h、j = 1, 2, \cdots, p;\ k、l = 1, 2, 3$$ 。

假定 $\boldsymbol{I}_{kl}(\boldsymbol{\vartheta}_{hj})$ 是有限的，并且矩阵 $\boldsymbol{I}(\boldsymbol{\vartheta}_{hj})$ 对所有 $\boldsymbol{\vartheta}_{hj} \in \omega_{hj}$ 都是正定的。

（5）对任意的 $\boldsymbol{\vartheta}_{hj} \in \omega_{hj}$，假定存在函数 $M_{klm}(y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$ 满足

$$\left|\frac{\partial^3}{\partial \vartheta_{hjk} \partial \vartheta_{hjl} \partial \vartheta_{hjm}} \ln\left[P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}^{\circ})\right]\right| \leqslant M_{klm}(y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$$

$$h、j = 1, 2, \cdots, p;\ k、l、m = 1, 2, 3$$

则将 $\tilde{\boldsymbol{\beta}}$ 代入由式（5.7）～式（5.9）组成的方程组依概率有解，并且 σ_{hj}、σ_h^2、σ_j^2 的边际似然估计 $\tilde{\sigma}_{hj}$、$\tilde{\sigma}_h^2$、$\tilde{\sigma}_j^2$ 是 $\sqrt{n}$ 相合的。

证明：第一步，证明存在性和相合性。

当 $\boldsymbol{\beta}$ 已知时，$P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})$ 是可识别的。令 $\boldsymbol{Q}_a$ 是以真值 $(\sigma_{hj}^{\circ}, \sigma_h^{\circ 2}, \sigma_j^{\circ 2})$ 为中心，半径为 a 的三维球体，对于任意充分小的 a，以概率 1 有

$$
\begin{aligned}
&E_{(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \beta^{\circ})}\left[\ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})\right] \leqslant \\
&E_{(\sigma_{hj}^{\circ}, \sigma_h^{\circ 2}, \sigma_j^{\circ 2}, \beta^{\circ})}\left[\ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})\right]
\end{aligned}
$$

对球面上所有的 $(\sigma_{hj}, \sigma_h^2, \sigma_j^2)$ 都成立。

上述结果证明的详细过程可以参看文献[69]。

由引理 5.4.2 得

$$
\begin{aligned}
&\frac{1}{n} l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) \to^{P} \\
&E_{(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \beta^{\circ})} \ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta}) \\
&\frac{1}{n} l_{hj}(\sigma_{hj}^{\circ}, \sigma_h^{\circ 2}, \sigma_j^{\circ 2}, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) \to^{P} \\
&E_{(\sigma_{hj}^{\circ}, \sigma_h^{\circ 2}, \sigma_j^{\circ 2}, \beta^{\circ})} \ln P_{hj}(y_h, y_j; \boldsymbol{X}_h, \boldsymbol{X}_j, \sigma_{hj}, \sigma_h^2, \sigma_j^2, \boldsymbol{\beta})
\end{aligned}
$$

所以对任意充分小的a，有

$$\frac{1}{n}l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j) \leqslant \frac{1}{n}l_{hj}(\overset{\circ}{\sigma}_{hj}, \overset{\circ}{\sigma}{}_h^2, \overset{\circ}{\sigma}{}_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)$$

对所有球面$\boldsymbol{Q}_a$上的$(\sigma_{hj}, \sigma_h^2, \sigma_j^2)$依概率成立。因此，依概率在$\boldsymbol{Q}_a$内部有局部最大值，将$\tilde{\boldsymbol{\beta}}$代入由式（5.7）～式（5.9）组成的方程组依概率有解，记其解为$\tilde{\boldsymbol{\vartheta}}_{hj} = (\tilde{\sigma}_{hj}, \tilde{\sigma}_h^2, \tilde{\sigma}_j^2)$。由$a$的任意性可知解是相合的。

第二步，证明$\sqrt{n}$的相合性。

将$\dfrac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}}$在$\overset{\circ}{\boldsymbol{\vartheta}}_{hj} = (\overset{\circ}{\sigma}_{hj}, \overset{\circ}{\sigma}{}_h^2, \overset{\circ}{\sigma}{}_j^2)$处进行 Taylor 展开可得

$$\begin{aligned}
&\frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}} \\
= &\frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}}\Big|_{\overset{\circ}{\vartheta}_{hj}} + \\
&\sum_{l=1}^{3}(\vartheta_{hjl} - \overset{\circ}{\vartheta}_{hjl})\frac{\partial^2 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\vartheta_{hjl}}\Big|_{\overset{\circ}{\vartheta}_{hj}} + \\
&\frac{1}{2}\sum_{l=1}^{3}\sum_{m=1}^{3}(\vartheta_{hjl} - \overset{\circ}{\vartheta}_{hjl})(\vartheta_{hjm} - \overset{\circ}{\vartheta}_{hjm})\frac{\partial l_{hj}^3(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\vartheta_{hjl}\vartheta_{hjm}}\Big|_{\vartheta_{hj}^*}
\end{aligned}$$

其中，$\boldsymbol{\vartheta}_{hj}^*$是$\boldsymbol{\vartheta}_{hj}$和$\overset{\circ}{\boldsymbol{\vartheta}}_{hj}$线段上的一点。由第一部分$\tilde{\boldsymbol{\vartheta}}_{hj}$依概率有解并是相合的，用$\tilde{\boldsymbol{\vartheta}}_{hj}$代替$\boldsymbol{\vartheta}_{hj}$得

$$\begin{aligned}
&\sqrt{n}\sum_{l=1}^{3}(\tilde{\vartheta}_{hjl}^{\circ} - \overset{\circ}{\vartheta}_{hjl})\left[\frac{1}{n}\frac{\partial^2 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\vartheta_{hjl}}\Big|_{\overset{\circ}{\vartheta}_{hj}} + \right. \\
&\left.\frac{1}{2n}\sum_{m=1}^{3}(\vartheta_{hjm} - \overset{\circ}{\vartheta}_{hjm})\frac{\partial l_{hj}^3(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\vartheta_{hjl}\vartheta_{hjm}}\Big|_{\vartheta_{hj}^*}\right] \\
= &-\frac{1}{\sqrt{n}}\frac{\partial l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}}\Big|_{\overset{\circ}{\vartheta}_{hj}}
\end{aligned}$$

记矩阵$\boldsymbol{A} = (a_{kl})$，其中元素$a_{kl}$为

$$a_{kl} = \frac{1}{n}\frac{\partial^2 l_{hj}(\sigma_{hj}, \sigma_h^2, \sigma_j^2, \tilde{\boldsymbol{\beta}}; y_h, y_j, \boldsymbol{X}_h, \boldsymbol{X}_j)}{\partial \vartheta_{hjk}\vartheta_{hjl}}\Big|_{\overset{\circ}{\vartheta}_{hj}} +$$

$$\frac{1}{2n}\sum_{m=1}^{3}(\vartheta_{hjm}-\vartheta_{hjm}^{\circ})\frac{\partial l_{hj}^{3}(\sigma_{hj},\sigma_h^2,\sigma_j^2,\tilde{\boldsymbol{\beta}};y_h,y_j,\boldsymbol{X}_h,\boldsymbol{X}_j)}{\partial\vartheta_{hjk}\vartheta_{hjl}\vartheta_{hjm}}|_{\vartheta_{hj}^{*}}$$

记 $\boldsymbol{T}=(t_1,t_2,t_3)^{\mathrm{T}}$ ，其中 $t_k=-\frac{1}{\sqrt{n}}\frac{\partial l_{hj}(\sigma_{hj},\sigma_h^2,\sigma_j^2,\tilde{\boldsymbol{\beta}};y_h,y_j,\boldsymbol{X}_h,\boldsymbol{X}_j)}{\partial\vartheta_{hjk}}|_{\vartheta_{hj}^{\circ}}$ $(k=1,2,3)$ 。

有

$$\sqrt{n}\boldsymbol{A}(\tilde{\boldsymbol{\vartheta}}_{hj}-\boldsymbol{\vartheta}_{hj}^{\circ})=\boldsymbol{T}$$

由估计的相合性可知矩阵 $\boldsymbol{A}$ 依概率是正定的，所以有

$$\sqrt{n}(\tilde{\boldsymbol{\vartheta}}_{hj}-\boldsymbol{\vartheta}_{hj}^{\circ})=\boldsymbol{A}^{-1}\boldsymbol{T}$$

记 $\boldsymbol{T}^{\circ}=(t_1^{\circ},t_2^{\circ},t_3^{\circ})^{\mathrm{T}}$，其中

$$t_k^{\circ}=-\frac{1}{\sqrt{n}}\frac{\partial l_{hj}(\sigma_{hj},\sigma_h^2,\sigma_j^2,\boldsymbol{\beta}^{\circ};y_h,y_j,\boldsymbol{X}_h,\boldsymbol{X}_j)}{\partial\vartheta_{hjk}}|_{\vartheta_{hj}^{\circ}},k=1,2,3$$

再由式（5.19）可知

$$\boldsymbol{A}=\boldsymbol{I}(\vartheta_{hj})+\boldsymbol{O}_p\left(\frac{1}{n}\right),T=\boldsymbol{T}^{\circ}+\boldsymbol{O}\left(\frac{1}{\sqrt{n}}\right)$$

而 $\boldsymbol{T}^{\circ}$ 是渐近正态的，渐近分布为 $N[\mathbf{0},\boldsymbol{I}(\vartheta_{hj})]$，所以有

$$\sqrt{n}(\tilde{\boldsymbol{\vartheta}}_{hj}-\boldsymbol{\vartheta}_{hj}^{\circ})=O_p(1)$$

得到 $\tilde{\boldsymbol{\vartheta}}_{hj}$ 是 $\sqrt{n}$ 相合的。

证毕。

由定理 5.4.1 可以得到 $\boldsymbol{\theta}$ 的渐近有效估计。

定理 5.4.1 在多重二元响应 Probit 模型中，参数 $\boldsymbol{\theta}\in\Theta$ ，并且 Θ 为一个开集。$\boldsymbol{I}(\boldsymbol{\theta})$ 是 $\boldsymbol{\theta}$ 的信息阵，$\boldsymbol{S}(\boldsymbol{\theta})$ 是得分函数。令参数真值为 $\boldsymbol{\theta}_0$，$\boldsymbol{I}(\boldsymbol{\theta}_0)$ 存在，并且 $\boldsymbol{I}(\boldsymbol{\theta}_0),\boldsymbol{S}(\boldsymbol{\theta}_0)$ 在 $\boldsymbol{\theta}_0$ 的邻域内连续。若模型满足引理 5.4.2 和引理 5.4.3 的条件，则式（5.20）中的线性估计 $\hat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的渐近有效估计。

$$\hat{\boldsymbol{\theta}}=\tilde{\boldsymbol{\theta}}+\hat{\boldsymbol{I}}(\tilde{\boldsymbol{\theta}})^{-1}\hat{\boldsymbol{S}}(\tilde{\boldsymbol{\theta}}) \tag{5.20}$$

并且有

$$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})\to_d N[\mathbf{0},\boldsymbol{S}(\boldsymbol{\theta})]$$

此定理可由文献[69]中相关内容得到。$\hat{\boldsymbol{S}}(\tilde{\boldsymbol{\theta}})$ 可由式（5.15）估计，$\hat{\boldsymbol{S}}(\tilde{\boldsymbol{\theta}})$ 中的元素可由式（5.16）估计，其中参数 $\boldsymbol{\theta}$ 由 $\tilde{\boldsymbol{\theta}}$ 代替。

上述方法当 $n\to\infty$ 时，在理论上只要一步迭代就可以得到有效估计，但在

有限样本时一次迭代可能得不到令人满意的结果，这时可以进一步迭代。由于在计算过程中只需要一步（或较少的几步）迭代。因此，在用 Monte-Carlo 方法计算信息阵 $\boldsymbol{I}(\tilde{\theta})$ 时可以增加抽取的随机数个数，以减小模拟方法产生的扰动。

5.5　模拟结果

本节利用模拟方法对上述方法进行研究，并与模拟极大似然估计进行比较。在模拟中参数真值取为 $\boldsymbol{\beta}=[1\ \ 2\ \ 0\ 1]$，协方差阵 $\boldsymbol{\Omega}$ 为

$$\boldsymbol{\Omega}=\begin{bmatrix}1 & 0.3 & 0.3 & 0.3\\ 0.3 & 2 & 0.5 & 0.5\\ 0.3 & 0.5 & 2 & 0.5\\ 0.3 & 0.5 & 0.5 & 2\end{bmatrix}$$

由于估计协方差阵元素时需要的样本容量较大，所以这里取样本容量为 n=500 和 n=1000 两种情形来模拟。

利用 Monte-Carlo 方法计算信息阵时，产生 10000 个随机数进行模拟计算。实验重复 100 次，估计结果如表 5.1 和表 5.2 所示。

同时利用模拟极大似然方法估计参数，并与二步方法进行比较。在模拟极大似然过程中仍然用 GHK 模拟方法，由于在求解过程中需要大量迭代，因此在计算高维积分时不可能产生太大数量的随机数。如果随机数较大，会使计算时间过长。这里与文献[16]中取相同的随机数个数，产生 500 个随机数计算高维积分。结果同样列入表 5.1 和表 5.2 中。

从模拟结果来看，通过迭代得到的估计确实能改进边际估计。但需要指出的是，渐近有效估计也是一种大样本方法，样本容量较小时效果不好。这是因为第二步的渐近有效估计必须要求第一步的边际估计是 $\sqrt{n}$ 相合的。

另外，在本章扰动项的协方差阵中，各元素的估计需要用到回归系数的边际估计，当样本容量较小时 σ_h^2、σ_j^2 和 σ_{hj} 的边际估计不准确，所以这里模拟时的样本量选择 n=500 和 n=1000。如果样本容量比较小，得到的边际估计不够准确，第二步迭代则不能改进边际估计。还有一点需要说明，利用模拟积分计算信息阵时要求产生随机数的数量较大，最好在 5000 次以上才能有较好的估计效果。因为本章中的渐近有效估计只需要一次迭代，所以产生随机数的数量可以取得很大，以减少模拟积分时随机扰动产生的误差。

表 5.1 $n = 500$ 时，参数估计的模拟结果

$n = 500$	β_1	β_2	β_3	β_4	σ_2^2	σ_3^2	σ_4^2	σ_{12}	σ_{13}	σ_{14}	σ_{23}	σ_{24}	σ_{34}
边际似然估计均值	1.0063	2.0070	0.0007	0.9929	1.9497	1.9717	1.9724	0.2871	0.3148	0.2953	0.4856	0.4858	0.4825
边际似然估计的标准差	0.1063	0.1255	0.1026	0.1053	0.2925	0.3062	0.3010	0.1451	0.1496	0.1529	0.2090	0.2052	0.2058
一次迭代估计的均值	0.9962	2.0057	0.0043	0.9967	1.9850	1.9714	1.9752	0.2922	0.2937	0.2866	0.4897	0.4890	0.4876
一次迭代估计的标准差	0.0872	0.0925	0.0882	0.0883	0.2921	0.2898	0.2862	0.1434	0.1383	0.1467	0.2001	0.1962	0.2056
二次迭代估计的均值	0.9967	2.0006	0.0044	0.9970	1.9620	1.9704	1.9929	0.2961	0.2882	0.2941	0.4860	0.4859	0.4786
二次迭代估计的标准差	0.0856	0.0898	0.0817	0.0831	0.2835	0.2821	0.2706	0.1419	0.1328	0.1462	0.1973	0.1918	0.1932
模拟似然估计	1.0025	1.9959	0.0078	0.9966	1.9905	1.9736	1.9777	0.3032	0.2882	0.2997	0.4931	0.4997	0.5103
模拟似然估计的标准差	0.0773	0.0824	0.0791	0.0726	0.3040	0.2958	0.3059	0.1348	0.1343	0.1327	0.1978	0.1935	0.1849

表 5.2　$n = 1000$ 时，参数估计的模拟结果

$n=1000$	β_1	β_2	β_3	β_4	σ_2^2	σ_3^2	σ_4^2	σ_{12}	σ_{13}	σ_{14}	σ_{23}	σ_{24}	σ_{34}
边际似然估计均值	0.9977	2.0038	−0.0026	1.0008	1.9837	1.9941	1.9878	0.2953	0.3089	0.2971	0.4884	0.4822	0.4948
边际似然估计的标准差	0.0599	0.0631	0.0592	0.0601	0.2362	0.2398	0.2451	0.1195	0.1214	0.1208	0.1688	0.1712	0.1695
一次迭代估计的均值	0.9970	1.9986	−0.0004	0.9965	2.0008	1.9766	1.9879	0.3003	0.2910	0.2968	0.4807	0.5012	0.5098
一次迭代估计的标准差	0.0584	0.0608	0.0585	0.0570	0.2236	0.2361	0.2353	0.1160	0.1140	0.1163	0.1599	0.1672	0.1687
二次迭代估计的均值	0.9978	1.9980	0.0017	0.9993	1.9932	1.9808	1.9792	0.2982	0.2978	0.2967	0.4894	0.5034	0.4940
二次迭代估计的标准差	0.0580	0.0589	0.0571	0.0567	0.2232	0.2286	0.2251	0.1113	0.1123	0.1160	0.1602	0.1608	0.1562
模拟似然估计	0.9990	2.0006	−0.0009	1.0021	1.9957	1.9821	1.9782	0.3060	0.3096	0.3071	0.4891	0.4983	0.5055
模拟似然估计的标准差	0.0593	0.0641	0.0596	0.0608	0.2411	0.2267	0.2258	0.1156	0.1177	0.1150	0.1627	0.1667	0.1616

第6章　多元秩-序Logit模型回归系数的极大似然估计

6.1　固定影响属性的多元秩-序模型

在式（1.1）和式（1.2）定义的有限因变量模型中，如果式（1.2）中的映射τ为

$$\boldsymbol{Y}=\tau(\boldsymbol{Y}^*)=(o_1, o_2, \cdots, o_p)$$

式中，$(o_1, o_2, \cdots, o_p)$是$(1, 2, \cdots, p)$的置换，满足$y_{o_1}^*\geqslant y_{o_2}^*\geqslant\cdots\geqslant y_{o_p}^*$则称模型为多元秩-序模型。如果潜在变量$\boldsymbol{Y}^*=(y_1^*, y_2^*, \cdots, y_p^*)$是各选择的效用值，则可以认为$\boldsymbol{Y}$是将各选项的效用由高到低的排序。

在统计调查分析中经常会建立多元秩-序模型对一些选择问题进行分析。在调查过程中，需要被调查者将选择对象进行排序，得到的排序数据包含更多信息，基于这种数据建立的模型会得到更准确的分析结果和选择概率预测。

另外，自变量矩阵$\boldsymbol{X}$的各行对应各个属性的效用，第2～4章讨论的方法都是针对自变量为随机的情形，即要求模型中各个属性的效用为随机的。但在有些应用中，属性的效应为几个固定值，不是随机变量，称之为固定效应模型。

例：学生选课问题的秩-序数据。为了配合某校英语公选课的开设，进行了一次统计调查分析。通过对三个班级的问卷调查获得了学生选课的秩-序数据。教师的相关属性如表6.1所示。

表6.1　教师的相关属性

教师编号	年龄	职称	学历	性别	近三年获得奖励情况
1	56	教授	本科	男	无
2	47	副教授	本科	女	市级
3	45	副教授	硕士	男	校级
4	40	副教授	硕士	男	校级

续表

教师编号	年龄	职称	学历	性别	近三年获得奖励情况
5	31	讲师	博士	男	无
6	27	助教	硕士	女	无
7	28	助教	硕士	女	无

以上是用文字表示的具体属性，不能用来直接进行统计分析，所以需要将其数量化，结果在列表 6.2 中。

表 6.2　教师的属性数据

教师编号	年龄（标准化）	职称	学历	性别	近三年获得奖励情况
1	1.5401	4	1	2	0
2	0.7178	3	1	1	2
3	0.5351	3	2	2	1
4	0.0783	3	2	2	1
5	−0.7439	2	3	2	0
6	−1.1094	1	2	1	0
7	−1.0180	1	2	1	0

在这个问题中，如果要调查教师的具体属性对选择结果的影响，则这些属性就是固定的，这些不变的属性对应的效用即为固定效用。对于这个问题我们可以建立多元秩-序 Logit 模型进行分析。

6.2　多元秩-序 Logit 模型的极大似然估计

6.2.1　多元秩-序 Logit 模型

设在一次调查中调查了 n 个个体，J 个被选择项中的第 j 个选择项的属性为 $\boldsymbol{x}_j=(x_1, x_2, \cdots, x_q)^{\mathrm{T}}$ $(j=1, 2, \cdots J)$。一般地，我们把第 n 个决策者选择第 j 个选择项的效用分解为两个部分，即

$$\boldsymbol{U}_{nj}=\boldsymbol{V}_{nj}+\boldsymbol{\varepsilon}_{nj} \tag{6.1}$$

通常认为固定效用 $\boldsymbol{V}_{nj}$ 和影响因素 $\boldsymbol{X}_j$ 之间满足线性关系，即

$$\boldsymbol{V}_{nj}=\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}$$

式中，$\boldsymbol{\beta}\in\mathbf{R}^q$，为 q 维列向量。$\boldsymbol{X}_j^{\mathrm{T}}\boldsymbol{\beta}$ 称为可观测到的效用，观测不到的效用归为扰动项 $\boldsymbol{\varepsilon}_{nj}$。假设 $\boldsymbol{\varepsilon}_{nj}$ 是 i.i.d.的随机变量，并且服从冈贝尔 I 型极值分布，

其密度和分布函数分别为

$$f(t)=\exp(-t)\exp\left[-\exp(-t)\right],\quad -\infty<t<+\infty$$

$$F(t)=\exp\left[-\exp(t)\right]$$

如果在多元秩-序模型中，效用式（6.1）中的随机效应服从上述分布，则称为多元秩-序 Logit 模型。

6.2.2　回归系数的极大似然估计

我们认为第 n 个决策者选择第 i 个选择项作为第一选择 $\Leftrightarrow$ 第 i 个选择项的效用 $\boldsymbol{U}_{ni}$ 大于其他所有选择所对应的效用 $\boldsymbol{U}_{nj}$， $j\neq i; j=1, 2, \cdots, i-1, i+1, \cdots, J$ 。这一事件的概率记为：$P_{ni}=\Pr\left(V_{ni}+\varepsilon_{ni}>V_{nj}+\varepsilon_{nj}, \forall j\neq i\right)=\Pr\left(\varepsilon_{nj}<\varepsilon_{ni}+V_{ni}-V_{nj}, \forall j\neq i\right)$。为了表示简单，我们将下标中的 n 省去。以下是 j 个选择项比其他 1，⋯，$j-1$ 个选择项效用大的 Logit 概率的推导。

$$\begin{aligned}&\Pr\left(U_j>U_1, U_j>U_2, \cdots, U_j>U_{j-1}\right)\\&=\Pr\left(\varepsilon_1<\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}-\boldsymbol{X}_1^{\mathrm{T}}\boldsymbol{\beta}+\varepsilon_j, \cdots, \varepsilon_{j-1}<\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}-\boldsymbol{x}_{j-1}^{\mathrm{T}}\boldsymbol{\beta}+\varepsilon_j\right)\\&=\int_{-\infty}^{\infty}\left\{\exp\left[-\sum_{1}^{j-1}\exp\left(-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}-\varepsilon_j\right)\right]\exp\left(-\varepsilon_j\right)\exp\left[-\exp\left(-\varepsilon_j\right)\right]\right\}\mathrm{d}\varepsilon_j\end{aligned}$$

令 $\exp\left(-\varepsilon_j\right)=t$

换元得

$$\begin{aligned}&\Pr\left(U_j>U_1, \cdots, U_j>U_{j-1}\right)\\&=\int_{+\infty}^{0}\left\{\exp\left[-t\sum_{1}^{j-1}\exp\left(-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)t\right]\exp(-t)\mathrm{d}t\right\}\end{aligned}$$

令 $v=-t$ ，即上式 $=\int_{-\infty}^{0}\exp(v)^{\sum_{i=1}^{j-1}\exp\left(-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}\exp(v)\mathrm{d}v$

令 $Z=\exp(v)$

可得

$$\begin{aligned}&\Pr\left(U_j>U_1, \cdots, U_j>U_{j-1}\right)\\&=\left[\frac{1}{\sum_{i=1}^{j-1}\exp\left(-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{x}_{j-1}^{\mathrm{T}}\boldsymbol{\beta}\right)+1}Z^{\sum_{i=1}^{j-1}\exp\left(-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)+1}\right]_0^1\end{aligned}$$

$$=\frac{1}{1+\sum_{i=1}^{j-1}\exp\left(-\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}+\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}=\frac{\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)}{\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)+\sum_{i}^{j-1}\exp\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}$$

$$=\frac{\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)}{\sum_{i=1}^{j}\exp\left(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}=\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)\cdot\left[\sum_{i=1}^{j}\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)\right]^{-1}$$

记 $\Pr\left(U_j>U_1,U_j>U_2,\cdots,U_j>U_{j-1}\right)=F_j\left[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_j;\boldsymbol{\beta}\right]$，即

$$F_j\left[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_j;\boldsymbol{\beta}\right]=\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)\left[\sum_{i=1}^{j}\exp\left(\boldsymbol{x}_j^{\mathrm{T}}\boldsymbol{\beta}\right)\right]^{-1}$$

推导完毕。

令选择顺序 $\boldsymbol{r}=\left(r_1,\cdots,r_M\right)$，满足 $U_{r_1}>U_{r_2}\cdots>U_{r_M}$，下面求秩-序选择概率，即求

$$\Pr\left(\boldsymbol{r},\boldsymbol{x};\boldsymbol{\beta}\right)=\Pr\left(U_{r_1}>U_{r_2}>\cdots>U_{r_M},\boldsymbol{x};\boldsymbol{\beta}\right)$$

我们先任取三个选择顺序 r_{m-2}、r_{m-1}、r_m $(2<m<M)$，则有

$$\begin{aligned}&\Pr\left(U_{r_{m-2}}>U_{r_{m-1}}>U_{r_m}\right)\\&=\Pr\left(U_{r_{m-2}}>U_{r_{m-1}},U_{r_{m-2}}>U_{r_m}\middle|U_{r_{m-1}}>U_{r_m}\right)\cdot\Pr\left(U_{r_{m-1}}>U_{r_m}\right)\\&=F_{m-2}\left[\boldsymbol{x}_{r_{m-2}},\boldsymbol{x}_{r_{m-1}},\boldsymbol{x}_{r_m};\boldsymbol{\beta}\right]\cdot F_{m-1}\left(\boldsymbol{x}_{r_{m-1}},\boldsymbol{x}_{r_m};\boldsymbol{\beta}\right)\end{aligned}$$

同理有

$$\Pr\left(\boldsymbol{r},\boldsymbol{x};\boldsymbol{\beta}\right)=\Pr\left(U_{r_1}>U_{r_2}>\cdots>U_{r_1}>U_{r_M}\middle|U_2>\cdots>U_{r_M}\right)\Pr\left(U_2>U_3>\cdots>U_{r_M}\right)$$

而

$$\Pr\left(U_2>U_3>\cdots>U_{r_M}\right)=P\left(U_{r_2}>U_{r_3}>\cdots>U_{r_2}>U_{r_M}\middle|U_{r_3}>U_{r_M}\right)$$

依此类推，由上面推导

$$\Pr\left(U_j>U_1>\cdots>U_j>U_{j-1}\right)=F_j\left[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_j;\boldsymbol{\beta}\right]$$

得

$$\Pr\left(\boldsymbol{r},\boldsymbol{x};\boldsymbol{\beta}\right)=\prod_{m=2}^{M}F_m\left[\boldsymbol{x}_{r_m},\cdots,\boldsymbol{x}_{r_{M-m+1}};\boldsymbol{\beta}\right]$$

式中，$F_m\left[\boldsymbol{x}_{r_m},\cdots,\boldsymbol{x}_{r_{M-m+1}};\boldsymbol{\beta}\right]$ 由式（1.5）给出。

可以得到对数似然函数，即

$$
\begin{aligned}
L(\beta) &= \sum_{n=1}^{N} \ln\left[\Pr\left(\boldsymbol{r}_n, \boldsymbol{x}_n; \boldsymbol{\beta}\right)\right] \\
&= \sum_{n=1}^{N} \ln \prod_{m=2}^{M} F_m\left[\boldsymbol{x}_{r_m}, \cdots, \boldsymbol{x}_{r_{M-m+1}}; \boldsymbol{\beta}\right]
\end{aligned}
$$

因为上式是从 m=2，…，M 开始的，我们将其变成从 m=1 开始，则有

$$
\begin{aligned}
L(\beta) &= \sum_{n=1}^{N} \ln\left[\Pr\left(\boldsymbol{r}_n, \boldsymbol{x}_n; \boldsymbol{\beta}\right)\right] \\
&= \sum_{n=1}^{N} \ln \prod_{m=2}^{M} F_m\left[\boldsymbol{x}_{r_m}, \cdots, \boldsymbol{x}_{r_{M-m+1}}; \boldsymbol{\beta}\right] \qquad (6.2) \\
&= \sum_{n=1}^{N} \sum_{m=1}^{M-1} \ln\left[F_{M-m+1}\left[\boldsymbol{x}_{r_M}, \cdots, \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right]
\end{aligned}
$$

则回归系数 $\boldsymbol{\beta}$ 的极大似然估计为式（6.2）达到最大值时 $\boldsymbol{\beta}$ 的解，或者对数似然方程

$$
\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0
$$

的解，其中

$$
\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{n=1}^{N} \sum_{m=1}^{M-1} \left\{ x_{r_m} - \left[\sum_{i=r_{m+1}}^{r_{M-1}} \exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta})\right]^{-1} \cdot \left[\sum_{i=r_{m+1}}^{r_{M-1}} \left(\exp(\boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}) \cdot \boldsymbol{x}_i\right)\right] \right\}
$$

6.2.3 模拟研究

本节利用模拟方法对上述方法进行研究。在模型中取固定影响因素 $\boldsymbol{X}$ 的值为一个 4×4 的矩阵，其值为

$$
\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} 2 & 1 & 1 & 0.5 \\ -1 & 1 & 1 & 3 \\ -0.5 & 1 & 0 & 2 \\ 2 & -1 & 1 & 3 \end{bmatrix}
$$

模型中有四个选择，矩阵中每行对应一个选择的固定影响因素。在模拟中，参数真值取 $\boldsymbol{\beta} = [1, 2, 0, 1]$， $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}$ 即为固定效应。以冈贝尔 I 型极值分布为随机影响因素，产生随机数 $\boldsymbol{\varepsilon}$，则 $\boldsymbol{V} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ 为模型中的各选择项的效用，根据效用最大化原理，比较效用 $\boldsymbol{V}$ 各分量的大小，可以得到选择的秩-序，即因变量 y。

由上述模型产生样本容量为 n=20，50，100，150 的随机数，估计回归系

数，即固定影响属性的权重，模拟重复 500 次来估计回归系数方向的均值及其标准差。将所得的结果列入表 6.3 中。

表 6.3　回归系数估计的均值及各分量的标准差（括号内为各分量的标准差）

n \ $\hat{\beta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$n=20$	1.1376	2.1337	−0.2749	1.0018
	（0.2320）	（0.2733）	（0.4955）	（0.1329）
$n=50$	1.0998	2.0154	−0.0277	1.0987
	（0.1691）	（0.1897）	（0.4047）	（0.1131）
$n=100$	0.9961	1.9970	−0.0016	1.0018
	（0.1030）	（0.1064）	（0.2476）	（0.0707）
$n=150$	0.9982	2.0002	−0.0026	0.9985
	（0.0837）	（0.0982）	（0.2029）	（0.0610）

从模拟结果来看，当样本容量为 $n=20$ 时，估计的误差较大，并且均值也有些偏差，说明估计不是无偏的。当样本容量增大时，估计的均值很好地收敛到参数真值，而且标准差也趋于 0，说明估计有渐近正态性和相合性。

当样本容量渐近增大时，估计的误差变小，当样本容量为 $n=100$ 时，估计的误差较小，估计结果已经很令人满意了。

下面的模拟是增加模型中的选项，将模型中的选项增加一个，也就是固定影响属性为下面的 5×4 矩阵，即

$$\boldsymbol{X}^{\mathrm{T}}=\begin{bmatrix} 2 & 1 & 1 & 0.5 \\ -1 & 1 & 1 & 3 \\ -0.5 & 1 & 0 & 2 \\ 2 & -1 & 1 & 3 \\ 1 & 1 & -1 & 1 \end{bmatrix}$$

回归系数的真值仍取 $\boldsymbol{\beta}=[1,2,0,1]$，产生样本容量为 n=20，50，100，150 的随机数，估计回归系数，模拟重复 500 次来估计回归系数方向的均值及其标准差。将所得的结果列入表 6.4 中。

表 6.4　回归系数估计的均值及各分量的标准差（括号内为各分量的标准差）

n \ $\hat{\beta}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$n=20$	1.1003	2.1312	−0.0195	1.0018
	（0.2267）	（0.2453）	（0.4443）	（0.1294）

续表

$\hat{\beta}$ \ n	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
$n=50$	1.0084	2.0232	−0.0017	0.9948
	（0.1652）	（0.1701）	（0.3609）	（0.0950）
$n=100$	0.9988	1.9980	0.0114	0.9985
	（0.0991）	（0.0946）	（0.2025）	（0.0535）
$n=150$	1.0108	2.0101	−0.0229	1.0056
	（0.0824）	（0.0831）	（0.1651）	（0.0460）

多元秩-序模型的选择项个数也会影响到参数估计，当模型中选择项较多时，其所包含的信息量也较多。所以当模型的选择项增加时，与前一个模拟比较，估计的精度也有所增加。这也说明了多元秩-序模型包含的信息量要比多项选择模型包含的信息量大，所以在设计调查过程中，如果能让被调查者对选择进行排序，则可以在保证相同估计准确度的前提下减少调查样的样本容量。

6.3　多元秩-序 Logit 模型的假设检验

前两节我们得到了多元秩-序 Logit 模型中回归系数的极大似然估计。为了得到似然函数的显式表达，假定扰动向量

$$\boldsymbol{\varepsilon}=\left(\varepsilon_1,\ \varepsilon_2,\ \cdots,\ \varepsilon_M\right)^{\mathrm{T}}$$

中的分量$\varepsilon_j\left(j=1,\ 2,\ \cdots,\ M\right)$的分布为 i.i.d.，并且服从极值分布。由于极值分布也是对称的，其形状与正态分布比较接近，所以在很多情形下这种假定是合理的。但是在得到模型中回归系数的极大似然估计之后，仍然需要检验模型的假定是否存在矛盾。模型中主要存在的假定错误为扰动向量的各分量不是 i.i.d.的，$\varepsilon_j\left(j=1,\ 2,\ \cdots,\ M\right)$的分布之间，或者是存在方差不同，或者是存在一定的相依关系。

在多元秩-序 Logit 模型中，考虑下面的假设检验：

原假设为：扰动项$\varepsilon_j\left(j=1,\ 2,\ \cdots,\ P\right)$i.i.d.且都服从极大值分布。

令$\hat{\boldsymbol{\beta}}$为多元秩-序 Logit 模型的极大似然估计，$\boldsymbol{\beta}$是回归系数的真值。在实际问题中回归系数的真值是未知的。由多元秩-序 Logit 模型的特点，可以构造秩-序 Logit 模型的部分似然函数，得到部分秩-序 Logit 模型的极大似然估计，

由后面的理论推导所得的极大似然估计的渐近正态性可以构造一个合理的检验统计量。

6.3.1　部分极大似然估计

多元秩-序 Logit 模型的似然函数为

$$
\begin{aligned}
L(\beta) &= \sum_{i=1}^{N} \ln\left[\Pr\left(r_n, \boldsymbol{X}; \boldsymbol{\beta}\right)\right] \\
&= \sum_{i=1}^{N}\sum_{m=1}^{M} \ln\left\{F_{M-m+1}\left[\boldsymbol{x}_{r_M} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right\} \\
&= \sum_{n=1}^{N}\sum_{m=1}^{P} \ln\left\{F_{M-m+1}\left[\boldsymbol{x}_{r_M} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right\} + \sum_{n=1}^{N}\sum_{m=p+1}^{M} \ln\left\{F_{M-m+1}\left[\boldsymbol{x}_{r_M} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right\}
\end{aligned}
$$

记

$$
L_p(\beta) = \sum_{n=1}^{N}\sum_{m=P+1}^{M} \ln\left\{F_{M-m+1}\left[\boldsymbol{x}_{r_M} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right\} \tag{6.3}
$$

$$
L_c(\beta) = \sum_{n=1}^{N}\sum_{m=P+1}^{M} \ln\left\{F_{M-m+1}\left[\boldsymbol{x}_{r_M} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right\} \tag{6.4}
$$

称式（6.3）与式（6.4）为部分似然函数。

由于多元秩-序 Logit 模型中扰动向量 $\boldsymbol{\varepsilon}$ 的各分量 i.i.d.，因此由式（6.3）与式（6.4）的部分似然函数也可以得到部分多元秩-序 Logit 模型中回归系数的极大似然估计。

记 $\hat{\boldsymbol{\beta}}_p$ 是使似然函数式（6.3）的 $L_p(\beta)$ 达到最大的部分似然估计，$\hat{\boldsymbol{\beta}}_c$ 是使似然函数式（6.4）的 $L_c(\beta)$ 达到最大的部分似然估计，即下面部分似然方程的解为

$$
\frac{\partial L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{P} \partial \ln\left\{F_{P-m+1}\left[\boldsymbol{x}_{r_P} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]\right\}}{\partial \boldsymbol{\beta}} = 0 \tag{6.5}
$$

$$
\frac{\partial L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{n=1}^{N}\sum_{m=1}^{P} \frac{\partial F_{M-m+1}\left\lfloor \boldsymbol{x}_{r_P} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right\rfloor}{\partial \boldsymbol{\beta}} \cdot \left[F_{P-m+1}\left(\boldsymbol{x}_{r_P} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right)\right]^{-1}
$$

其中

$$
\frac{\partial F_{M-m+1}\left[\boldsymbol{x}_{r_P} \cdots \boldsymbol{x}_{r_m}; \boldsymbol{\beta}\right]}{\partial \boldsymbol{\beta}}
$$

$$
=\frac{\partial\left\{\exp(\boldsymbol{x}_{rm}^{\mathrm{T}}\boldsymbol{\beta})\left[\sum_{i=r_{M+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]\right\}^{-1}}{\partial\boldsymbol{\beta}}
$$

$$
=\exp(\boldsymbol{x}_{r_M}^{\mathrm{T}}\boldsymbol{\beta})\cdot\left[\sum_{i=r_{M+1}}^{r_P}\exp(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\cdot\boldsymbol{x}_{r_m}-
$$

$$
\exp(\boldsymbol{x}_{rm}^{\mathrm{T}}\boldsymbol{\beta})\cdot\left[\sum_{i=r_{m+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]^{-2}\sum_{i=r_{m+1}}^{r_P}\left[\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\cdot\boldsymbol{x}_i\right]
$$

$$
=\left[\sum_{i=r_m}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\cdot\left\{\exp(\boldsymbol{x}_{r_m}^{\mathrm{T}}\boldsymbol{\beta})\cdot\left(\boldsymbol{x}_{r_m}-\left[\sum_{i=r_{m+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\sum_{i=r_{m+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{x}_i\right)\right]\right\}
$$

所以

$$
\frac{\partial L_p(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\sum_{n=1}^{N}\sum_{m=1}^{P}\left[\sum_{i=r_m}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\cdot\left\{\exp(\boldsymbol{x}_{r_m}^{\mathrm{T}}\boldsymbol{\beta})\cdot\left[\boldsymbol{x}_{r_m}-\left(\sum_{i=r_{m+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right)^{-1}\cdot\right.\right.
$$

$$
\left.\left.\sum_{i=r_{m+1}}^{r_P}\left(\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\cdot\boldsymbol{x}_i\right)\right]\right\}\cdot\left\{\exp(\boldsymbol{x}_{rm}^{\mathrm{T}}\boldsymbol{\beta})\cdot\left[\sum_{i=r_{m+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\right\}^{-1} \tag{6.6}
$$

$$
=\sum_{n=1}^{N}\sum_{m=1}^{P}\left\{\boldsymbol{x}_{r_m}-\left[\sum_{i=r_{m+1}}^{r_P}\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\cdot\left[\sum_{i=r_{m+1}}^{r_P}\left(\exp(\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta})\cdot\boldsymbol{x}_i\right)\right]\right\}
$$

实际上 $\hat{\boldsymbol{\beta}}_p$ 是由前 P 个选择项构成的部分秩-序模型的部分极大似然估计；同理，也可以由后 $M-P$ 个选择项构成部分秩-序模型得到部分极大似然估计 $\hat{\boldsymbol{\beta}}_c$，本节主要应用 $\hat{\boldsymbol{\beta}}_p$ 来构造检验。

6.3.2 极大似然估计的渐近正态性及相关结论

在构造检验统计量时，要应用到极大似然估计的渐近正态性及有关的结论，这里给出简要介绍。

定理 6.3.1 设 $x_1\cdots x_n$ 是来自 $p(x,\boldsymbol{\theta})$ 的一个样本，参数空间 $\varTheta$ 为一个开区间，若 $\ln P(x;\boldsymbol{\theta})$ 在 $\varTheta$ 上可微，并设 $p(x;\boldsymbol{\theta})$ 是可识别的（$\forall\boldsymbol{\theta}\neq\boldsymbol{\theta}^{\mathrm{T}}, \{x: p(x;\boldsymbol{\theta})\neq p(x;\boldsymbol{\theta}^{\mathrm{T}})\}$ 不是零测集）。

概率密度 $p(x;\boldsymbol{\theta})$，$\boldsymbol{\theta}\in\varTheta$ 满足：

（1）在参数真值$\boldsymbol{\theta}_0$的邻域内，$\dfrac{\partial \ln p(x,\ \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$，$\dfrac{\partial^2 \ln p(x,\ \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}$ $\dfrac{\partial^3 \ln p(x,\ \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^3}$对所有$x$都存在。

（2）在参数真值的邻域内$\left|\dfrac{\partial^3 \ln p(x,\ \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^3}\right| \leqslant \boldsymbol{H}(x)$，且$E\boldsymbol{H}(x) < \infty$。

（3）在参数真值$\boldsymbol{\theta}_0$处

$$E_{\theta_0}\left[\frac{p'(x,\ \boldsymbol{\theta}_0)}{p(x,\ \boldsymbol{\theta})}\right] = 0$$

$$E_{\theta_0}\left[\frac{p''(x,\ \boldsymbol{\theta}_0)}{p(x,\ \boldsymbol{\theta})}\right] = 0$$

$$I(\boldsymbol{\theta}_0) = E_{\theta}\left[\frac{p'(x,\ \boldsymbol{\theta}_0)}{p(x,\ \boldsymbol{\theta}_0)}\right]^2 > 0$$

式中，撇号表示对$\boldsymbol{\theta}$的微分，记$\hat{\boldsymbol{\theta}}_n$为$n \to \infty$时似然方程的相合解。则

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{L} N\left[\mathbf{0},\ \boldsymbol{I}^{-1}(\boldsymbol{\theta}_0)\right] \tag{6.7}$$

若$\boldsymbol{\theta}$为K维向量时，Θ是M维欧氏空间$\mathbf{R}^K$中一个含有内点的集合：

$$I(\boldsymbol{\theta}) = [I_{ij}(\boldsymbol{\theta})]_{K \times K} > 0$$

$$I_{ij}(\boldsymbol{\theta}) = \int_x \left[\frac{\partial \ln p(x;\ \boldsymbol{\theta})}{\partial \theta_i}\right] \cdot \left[\frac{\partial \ln p(x;\ \boldsymbol{\theta})}{\partial \theta_j}\right] \cdot p(x;\ \boldsymbol{\theta})\mathrm{d}u(x)$$

$$= E_{\theta}\left[\frac{\partial \ln p(x;\ \boldsymbol{\theta})}{\partial \theta_i}\right] \cdot \left[\frac{\partial \ln p(x;\ \boldsymbol{\theta})}{\partial \theta_j}\right]$$

$$i、j = 1,\ \cdots,\ K$$

则有

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{L} N\left[0,\ \boldsymbol{I}^{-1}\left(\boldsymbol{\theta}_0\right)\right]$$

由定理 6.3.1 可知，多元秩-序模型的极大似然估计$\hat{\beta}$是渐近正态的，有

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \beta_0\right) \xrightarrow{L} N\left[0,\ \boldsymbol{I}^{-1}\left(\beta_0\right)\right] \tag{6.8}$$

$\boldsymbol{I}(\boldsymbol{\beta})$为一个$q \times q$对称矩阵，$q$是回归系数向量$\boldsymbol{\beta}$的维数，则有

$$\boldsymbol{I}_{ij}\left(\boldsymbol{\beta}\right) = E_{\beta}\left(\frac{\partial L\left(\boldsymbol{\beta}\right)}{\partial \beta_i} \cdot \frac{\partial L\left(\boldsymbol{\beta}\right)}{\partial \beta_j}\right),\quad i、j = 1,\ \cdots,\ q$$

其中

$$\frac{\partial L(\boldsymbol{\beta})}{\partial \beta_i}=\sum_{n=1}^{N}\sum_{m=1}^{M-1}\left\{x_{r_m},i-\left[\sum_{k=r_{m+1}}^{r_M}\exp\left(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\beta}\right)\right]^{-1}\cdot\sum_{k=r_{m+1}}^{r_M}\left[\exp(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\beta})\cdot x_k\right]\right.$$

$$i=1,\ \cdots,\ k$$

部分似然估计也是渐近正态的，有

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}\right)\xrightarrow{L}N\left[0,\boldsymbol{I}_p^{-1}\left(\boldsymbol{\beta}\right)\right] \tag{6.9}$$

$$I_{P(ij)}\left(\boldsymbol{\beta}\right)=E_{\beta}\left(\frac{\partial L_P\left(\boldsymbol{\beta}\right)}{\partial \beta_i}\cdot\frac{\partial L\left(\boldsymbol{\beta}\right)}{\partial \beta_i}\right)$$

其中，$\dfrac{\partial L_P\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}}$由式（6.6）给出。

综合式（6.8）和式（6.9）有

$$\sqrt{n}\begin{pmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \hat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}\end{pmatrix}\xrightarrow{L}N\left(\boldsymbol{0},\boldsymbol{H}_1\right)$$

其中

$$\boldsymbol{H}_1=\begin{pmatrix}\boldsymbol{I}\left(\boldsymbol{\beta}\right) & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_P\left(\boldsymbol{\beta}\right)\end{pmatrix}$$

令$\boldsymbol{Q}=\begin{pmatrix}\boldsymbol{E}_q\\ -\boldsymbol{E}_q\end{pmatrix}$，其中$\boldsymbol{E}_q$为$q$阶单位阵，

则$\sqrt{n}\boldsymbol{Q}^{\mathrm{T}}\begin{pmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \hat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}\end{pmatrix}\xrightarrow{L}N\left(\boldsymbol{0},\boldsymbol{H}_2\right)$。

其中

$$\boldsymbol{H}_2=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{Q}=\left[\boldsymbol{E}_q\ -\boldsymbol{E}_q\right]\begin{bmatrix}\boldsymbol{I}^{-1}(\boldsymbol{\beta}) & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{I}_P^{-1}\left(\boldsymbol{\beta}\right)\end{bmatrix}\begin{bmatrix}\boldsymbol{E}_q\\ -\boldsymbol{E}_q\end{bmatrix}=\boldsymbol{I}^{-1}\left(\boldsymbol{\beta}\right)+\boldsymbol{I}_P^{-1}\left(\boldsymbol{\beta}\right)$$

由

$$\boldsymbol{Q}^{\mathrm{T}}\begin{pmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \hat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}\end{pmatrix}=\left[\boldsymbol{E}_q,-\boldsymbol{E}_q\right]\begin{bmatrix}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \hat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}\end{bmatrix}=\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}_p-\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_p$$

得

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_p\right)\xrightarrow{L}N\left(\boldsymbol{0},\boldsymbol{H}_2\right)$$

以上就是极大似然估计与部分似然估计，在模型的假设成立的前提下的渐

近分布。利用这个渐近分布可以对模型进行检验。

6.3.3　检验统计量

H_0：检验的原假设为

扰动项 $\varepsilon_j\left(j=1, 2, \cdots, P\right)$ 服从 i.i.d.且都服从极大值分布。

令检验统计量 $T=n(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_p)^{\mathrm{T}}\boldsymbol{H}_2^{-1}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_p)$，由前面给出的 $\sqrt{n}(\hat{\boldsymbol{\beta}}-\hat{\boldsymbol{\beta}}_p)$ 的渐近正态性可知，当原假设成立时有

$$T\to x^2(q)$$

式中，q 是卡方分布的自由度，且 q 是回归系数 $\boldsymbol{\beta}$ 的维数。

需要指出的是，$\boldsymbol{H}_2$ 是未知的，但是可以由数据估计得到。下面给出 $\boldsymbol{H}_2$ 的相合估计。

$\boldsymbol{H}_2$ 中 $\boldsymbol{I}_P(\boldsymbol{\beta})$ 的估计为

$$\boldsymbol{I}_P\left(\hat{\boldsymbol{\beta}}_p\right)=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{\partial\ln\Pr\left(\boldsymbol{r}_{np}\boldsymbol{x};\boldsymbol{\beta}\right)}{\partial\beta_i}\cdot\frac{\partial\ln\Pr\left(\boldsymbol{r}_{np}\boldsymbol{x};\boldsymbol{\beta}\right)}{\partial\beta_j}\right)\Bigg|_{\beta=\hat{\beta}_p},\quad i、j=1,\ \cdots,\ q$$

式中，$\Pr\left(\boldsymbol{r}_{np}\boldsymbol{x};\boldsymbol{\beta}\right)$ 为第一部分的 $\prod\limits_{m=2}^{M}F_m\left[\boldsymbol{x}_{r_m},\ \cdots,\boldsymbol{x}_{r_{M-m+1}};\boldsymbol{\beta}\right]$ 式，而

$$\frac{\partial\ln\Pr(\boldsymbol{r}_{np},\boldsymbol{x};\boldsymbol{\beta})}{\partial\beta_i}=\sum_{m=1}^{P}\left\{\boldsymbol{x}_{r_{mi}}-\left[\sum_{k=r_{m+1}}^{r_P}(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\left[\sum_{k=r_{m+1}}^{r_P}\left(\exp(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{x}_{ki}\right)\right]\right\}$$

式中，$\boldsymbol{r}_{np}$ 表示提取数据中所对应的前 P+1 顺序选择，$\boldsymbol{x}_{r_{mi}}$ 与 $\boldsymbol{x}_{ki}$ 分别表示向量 $\boldsymbol{x}_{r_m}$ 和 $\boldsymbol{x}_k$ 中的第 i 个分量。

$\boldsymbol{I}(\boldsymbol{\beta})$ 的估计为

$$\boldsymbol{I}(\hat{\boldsymbol{\beta}})=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{\partial\ln\Pr(\boldsymbol{r}_n,\boldsymbol{x};\boldsymbol{\beta})}{\partial\beta_i}\cdot\frac{\partial\ln\Pr(\boldsymbol{r}_n,\boldsymbol{x};\boldsymbol{\beta})}{\partial\beta_j}\right)\Bigg|_{\beta=\hat{\beta}},\quad i、j=1,\ \cdots,\ q$$

$$\frac{\partial\ln\Pr(\boldsymbol{r}_{np},\boldsymbol{x};\boldsymbol{\beta})}{\partial\beta_i}=\sum_{m=1}^{M-1}\left\{\boldsymbol{x}_{r_{mi}}-\left[\sum_{k=r_{m+1}}^{r_M}(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\beta})\right]^{-1}\left[\sum_{k=r_{m+1}}^{r_M}\left[\exp(\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{x}_{ki}\right]\right]\right\}$$

根据检验统计量 T 的渐近分布可以得到检验式（6.10）的拒绝域为

$$\left\{\left(r,x\right)=T\geqslant x_{1-\alpha}^2(q)\right\}$$

式中，$x_{1-\alpha}^2(q)$ 是自由度为 q 下 $1-\alpha$ 分位数，$1-\alpha$ 为给定的显著性水平，通常取 5%或 10%。

上面给出的检验是合理的，我们主要检验$\varepsilon_j\left(j=1,\cdots,M\right)$是否 i.i.d.。当原假设成立时，$\hat{\beta}$与$\hat{\beta}_p$都是相合估计，所以$\hat{\beta}$与$\hat{\beta}_p$的差异较小。当数据显著地拒绝原假设时，说明$\hat{\beta}$与$\hat{\beta}_p$差异较大，而$\hat{\beta}_p$是由部分秩-序模型得到的，就说明$\varepsilon_j$中存在非 i.i.d.变量。

6.3.4 假设检验的模拟研究

在以下的模拟研究中，令固定影响属性为5×4的矩阵，即

$$\boldsymbol{X}^{\mathrm{T}}=\begin{bmatrix}2&1&1&0.5\\-1&1&1&3\\-0.5&1&0&2\\2&-1&1&3\\1&1&-1&1\end{bmatrix}$$

回归系数的真值仍取$\beta=[1,2,0,1]$，产生样本容量为 n=20，50，100，150 的随机数，然后进行假设检验，求部分极大似然时用前四个选项。模拟重复 1000 次检验犯第一类错误的频率。分别模拟显著性水平为 0.05 和 0.1 的检验，将所得的结果列入表 6.5 中。

表 6.5 检验犯第一类错误的频率

检验的水平 \ 样本容量	$n=20$	$n=50$	$n=100$	$n=150$
$\alpha=0.05$	0.077	0.061	0.052	0.049
$\alpha=0.1$	0.138	0.121	0.105	0.098

从模拟结果来看，当样本容量较小时，犯第一类错误的频率较高，而当样本容量增大时，检验犯第一错误的频率逐渐收敛到名义水平，说明回归系数的极大似然估计与部分极大似然估计都是渐近正态的。

6.4 实例分析

仍然回到本章开始提到的实际例子，我们对三个班级 108 名学生选择英语课教师的情况（英语课教师的相关属性同表 6.1）进行问卷调查。剔除 5 份无效问卷，获得了 103 名学生选教师的秩-序数据，如表 6.6 所示。

表 6.6　学生选教师的秩-序数据

教师编号	排序 1	排序 2	排序 3	排序 4	排序 5	排序 6	排序 7
1	5	3	16	17	25	22	15
2	26	20	19	17	8	7	6
3	19	29	20	14	11	6	4
4	28	22	22	10	14	4	3
5	20	19	17	22	12	8	5
6	4	3	4	14	15	21	42
7	1	7	5	9	18	35	28

表 6.1 中教师的属性除年龄外均为文字表达。用文字表示的属性，不能用其来直接进行统计分析，必须用相应的属性变量表示。其中职称、学历、获奖情况是有序的属性变量，这里分别用 0、1、2、…，由低到高表示这些属性。职称属性具体表示方法为助教、讲师、副教授、教授分别用 1、2、3、4 表示；学历属性具体表示方法为本科、硕士、博士分别用 1、2、3 表示。获奖属性具体表示方法为无获奖用 0 表示，获校级奖用 1 表示，获市级奖用 2 表示，获省级奖用 3 表示，获国家级奖用 4 表示。性别虽然不是有序的属性变量，没有高低之分，但在实际中学生选择的时候可能会考虑教师的性别，因此建模时也要考虑到性别可能对选择结果产生的影响。对性别属性具体表示方法为女教师用 1 表示，男教师用 2 表示。注意到年龄属性的数值比其他属性的数值大得多，会给分析带来不便，所以必须要将年龄数据标准化，步骤如下：

令 x_1 为年龄，$\overline{x}_1$ 为年龄数据的平均值，σ 为年龄数据的标准差，x_1^* 为标准化后的年龄，计算公式为

$$x_1^* = (x_1 - \overline{x}) / \sigma$$

经过量化之后的教师属性数据同表 6.2。

将调查数据代入得到效用权 β 的极大似然估计为

$$\hat{\beta} = \begin{bmatrix} -0.1139 & 0.8397 & 1.1011 & -0.5501 & 0.5794 \end{bmatrix}^{\mathrm{T}}$$

在进行检验时，选定 $p = 4$，得到检验统计量 T 的观察值为 7.0826，取检验的水平为 0.05，查表 $\chi^2_{0.05}(5) = 11.070$。因为 $T < \chi^2_{0.05}(5)$，所以不能拒绝原假设，可以认为式（6.10）的假定是可以接受的。

对应于年龄项的效用权为−0.1139，说明学生偏好于比较年轻的教师，由于数据中年龄数据进行了标准化，且该效用权的绝对值较小，所以年龄的差异对选择的影响不大。在实际选择中，年龄较大的老师的选择顺序并不靠后，说明

老教师赢得学生的青睐不是依靠年龄，而是依靠较高的职称（说明其具有深厚的学识和丰富的教学经验）、教学中获得的丰硕成果（获奖情况）等因素。在后面几项中，职称、学历、获奖情况等都有较大的影响，说明学生比较看中教师的综合能力。同时我们还发现性别对选课也有较大的影响，学生比较偏好于女教师，这与学科有关系，因为这是选择英语公选课教师的调查，在语言的组织方面，女教师可能比男教师更有优势。

用量化分析方法来研究管理问题正成为研究的热点，但是行为方式的不确定性给量化分析带来了很多困难。本章尝试利用计量经济学中的模型量化来分析教学管理方面的问题，得出的结果可以应用于对选择概率的预测。

参考文献

[1] Hajivassiliou V. Simulation Estimation Methods for Limited Dependent Variable Models[A]. Handbook of Statistics[C]. Amsterdam: North-Holland, 1993, 11: 519-543.

[2] Keane M P. A note on identification in the multinomial probit model[J]. Journal of Business and Economic Statistics, 1992, 10(2): 193-200.

[3] McFadden D. Conditional Logit Analysis of Qualitative Choice Behavior[A]. Frontiers in Econometrics[C]. New York: Academic Press，1973: 105-142.

[4] Heckman J. Dynamic Discrete Models Structural Analysis of Discrete Data with Econometric Applications[M]. Cambridge, MA: MIT Press, 1981.

[5] Bolduc D. Generalized Autoregressive Errors in the Multinomial Probit Model[J]. Transportation Reseach B-Methodological, 1992, 26(2): 155-170.

[6] Bolduc D, Kac M. Multinomial Probit Models with Factor-based Autoregressive Errors: A Computationally Efficient Estimation Approach[R]. Econpapers: University Laval, 1991.

[7] Hausman J, Wise D. A Conditional Probit Model for Qualitative Choice: Discrete Decisions-Recognizing Interdependentce and Heterogeneous Preferences[J]. Econometrica, 1978, 46(2): 403-426.

[8] Lerman S, Manski C. On the use of simulated frequencies to approximate choice probabilities [A]. Structural analysis of disrete data with Econometric applications[C]. Cambridge, MA: MIT Press，1981: 305-319.

[9] Hendry D. Monte Carlo Experimentation in Econometrics[A]. Handbook of Econometrics[C], Amsterdam：North Hollard，1984, 2: 937-976.

[10] Hajivassiliou V, McFadden D. The method of simulated scores with application to models of external debt crises[R]. Cowles Foundation Disscussion Paper: No.967, 1990.

[11] Kean M. A computationally efficient practical simulation estimator for panel data, with applications to estimating temporal dependence in employment and wages[R]. Mimeo: University of Minnesota, 1990.

[12] 高惠璇. 统计计算[M]. 北京：北京大学出版社, 1995.

[13] Borsch-Supan A, Hajivassiliou V. Smooth Unbias Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models[J]. journal of

Econometrics, 1993, 58(3): 347-368.

[14] Severini T A. Seeking efficient data augmentation schemes via conditional and marginal augmenttation[J]. Biometrika, 1999, 86(2): 301-320.

[15] Newey W K, McFadden D. Large sample estimation and hypothesis testing[A]. Handbook of econometrics[C]，Amsterdam: North Holland: 1994, 4: 2113-2148.

[16] Hajivassiliou V, McFadden D, Ruud P. Simulation of multivariate normal rectangel probabilities and their derivatives theoretical and computational results[J]. Journal of Econometrics, 1996, 72(1): 85-134.

[17] Hajivassiliou V, McFadden D. The method of simulated scores for the estimation of LDV models[J]. Econometrica, 1998, 66(3): 863-896.

[18] Natarajan R, McCulloch E, Kiefer N. A Monte Carlo EM method for estimating multinomial probit models[J]. Computational Statistics and Data Analysis，2000, 34(1): 33-50.

[19] 刘金燕，徐兴忠．多周期 Probit 模型中 MLE 的存在性[J]．应用数学, 2004, 17(增): 85-89.

[20] 孙立敏，徐兴忠．多元秩-序 Probit 模型中的 MLE 的存在性[J]．应用数学，2004, 17(增): 115-119.

[21] 赵江涛，徐兴忠．多项 Probit 模型参数的极大似然估计[J]．应用数学,2004 17(增): 90-93.

[22] Nobile A. A hybrid Markov chain for the Bayesian analysis of the multinomial probit model [J]. Statistics and Computing, 1998, 8(3)：229-242.

[23] Nobile A. Comment: Bayesian multinomial probit models with normalization constraint[J]. Journal of Econometrics, 2000, 99(2)：335-345.

[24] McCulloch R, Rossi P. An exact likelihood analysis of the multinomial probit model[J]. Journal of Econometrics, 1994, 64(1): 207-240.

[25] McCulloch R, Rossi P. Reply to Nobile[J]. Journal of Econometrics, 2000, 99(2): 347-348.

[26] McCulloch R, Polson N G, Rossi P. A Bayesian analysis of the multinomial probit model with fully identified parameters[J]. Journal of Econometrics, 2000,99(1): 173-193.

[27] McFadden D. A method of simulated moments for estimation of discrete response models without numerical integration[J]. Econometrica, 1989, 57(5):　995-1026.

[28] Pakes A, Pollard D. The Asymptotic Distribution of Simulation Estimators[J]. Econometrica，1989, 57(5)：1027-1057.

[29] Liu J S, Wong W H，Kong A.Covariance structure of the Gibbs sampler with applications to comparisons of estimators and augmentation schemes[J]. Biometrika，1994, 81(1)：27-40.

[30] Maddala G. Limited-dependent and Qualitative Variables in Econometrics[M]. Cambridge：

Cambridge University Press, 1983.

[31] McFadden D, lye W, Train K. An application of diagnostic tests for the independence from irrelevant alternatives property of the multinomial logit model[J]. Transportation research record, 1978, 637：39-45.

[32] Ruud P. Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomialdiscrete choice models[J]. Econometrica, 1983, 51(1): 225-228.

[33] Ruud P. Consistent estimation of limited dependent variable models despite misspecification of distribution[J]. Journal of Econometrics, 1986, 32(1): 157-187.

[34] Duan N, Li K C. Slicing regression: a link-free regression method[J]. The Annals of Statistics, 1991, 19(2)：505-530.

[35] Li K C. Sliced Inverse Regression for Dimmension Reduction(with discussion)[J]. Jounnal of the American Statistical Association, 1991, 86(2): 316-342.

[36]Kai T F，Samuel K, Wang K N. Symmetric Multivariate and Related Distribution[M]. London, UK: Chapman and Hall, 1990.

[37] Cambanis S, Huang S, Simons G. On the theory of elliptically contoured distribution[J]. Journal of Multivariate Analysis, 1981, 11(3)：368-385.

[38] Eton M L. A characterization of sherical distributions[J]. Journal of Multivariate Analysis, 1986, 20(2): 272-276.

[39] Hall P, Li K C. On Almost linearity of low dimensional projections from high dimensional data[J]. The Annals of Statistics, 1993, 21(2): 867-889.

[40] Li K C. Nonlinear confounding in high-dimensional regression[J]. The Annals of Statistics, 1997, 25(2): 577-612.

[41] Aragon Y, Satrcco J. Sliced inverse regression(S.I.R.): An appraisal of small sample alternatives to slicing[J]. Computational Statistics, 1996, 24(24), 107-118.

[42] Schott J R. Determining the Dimensionality in SIR[J]. Journal of the AmericanStatistical Association. 1994, 89(425): 141-148.

[43] Zhi Y, Robert E. Using the Bootstrap to Select One of a New Class of Dimension Reduction Methods[J]. Journal of the AmericanStatistical Association, 2003, 98(464): 968-979.

[44] Li K C. Data visualization with SIR: a transformation based projection pursuit method[R]. UCLA: Statistical Series, 24, 1989.

[45] Hsing T, Carroll R. An asymptotic theory for sliced inverse regression[J]. The Annals of

Statistics, 1992, 20(2): 1040-1061.

[46] Zhu L X, Ng K W. Asymptotic of sliced inverse regression[J]. Statistica Sinica, 1995, 5(2): 727-736.

[47] Zhu L X, Fang K T. Asymptotic for Kernel Estimate of Sliced Inverse Regression[J]. The Annals of Statistics, 1995, 24(3): 1053-1068.

[48] Fung W, He X, Liu L, et al. Dimension reduction based on canonical correlation[J]. Statistica Sinica. 2002, 12(12): 1093-1113.

[49] Cook R D. Graphics for regressions with a binary response[J]. Journal of the AmericanStatistical Association, 1996, 91(435): 983-992.

[50] Cook R D, Lee H. Dimension reduction in binary responseregression[J]. Journal of the AmericanStatistical Association, 1999, 94(448): 1187-1200.

[51] Hsing T. Nearest neighbor inverse regression[J]. The annals of Statistics. 1999, 27(2), 697-731.

[52] Cook R D, Weisberg S. Comment on Sliced inverse regression for dimension reduction (with discussion)[J]. Journal of the AmericanStatistical Association, 1991, 86(414): 328-332.

[53] Li K C, Aragon Y, Shedden K, et al. Dimension reduction for Multivariate response data[J]. Journal of the AmericanStatistical Association, 2003, 98(461), 99-109.

[54] Naik P, Tsai C L. Constrained inverse regression for incorporating prior information[J]. Journal of the AmericanStatistical Association, 2005, 100(469): 204-211.

[55] Li K C, Wang J L, Chen C H. Dimension reduction for censored regression data[J]. The Annals of Statistics, 1999, 27(1): 1-23.

[56] Raymond J C, Li K C. Measurement error rgression with unknown link: Dimension reduction and dara viualization[J]. Journal of the AmericanStatistical Association, 1992, 87(420), 1040-1050.

[57] Schwarz G. Estimating the dimension of a model[J]. The Annals of Statistics, 1978, 6(2), 461-464.

[58] Prasad A N, Tsai C L. Constrained inverse regression for incorporating prior information[J]. Journal of the AmericanStatistical Association, 2005, 100(469): 204-211.

[59] Ferre L. Dimension Choice for Sliced Inverse Regression Based on Rank[J]. Comptes Rendus de l Académie des Sciences - Ser, 1996 , 2 (4): 95-108.

[60] Li L, Cook R D, Christopher J N. Model-free variable selection[J]. Journal of the Royal Statistical Society B, 2005, 67(2): 285-299.

[61] Bura E, Cook R D. Estimating the structrual dimension of regressions via parametric inverse regression[J]. Journal of the Royal Statistical Society B, 2001, 63(2): 393-410.

[62] Cook R D, Messan C S. A Model free test for reduced rank in Multivariate regression[J]. Journal of the AmericanStatistical Association, 2003, 98(462): 340-351.

[63] Saracco J. Sliced inverse regression under linear constraints[J]. Communications in Statistics-Theory and Methods, 1999, 28(10): 2367-2393.

[64] Saracoo J. An asymptotic theory for sliced Inverse Regression[J]. Communications in Statistics,Theory and methods, 1997, 26(9): 2141-2171.

[65] Efron B. Bootstrap methods: Another look at the jackknife[J]. The Annals of Statistics, 1979, 7(1): 1-26.

[66] Jun Shao, Dongsheng Tu. The Jackknife and Bootstrap[M]. New York: Springer-Verlag, 1995.

[67] Tyler D E. Asymptotic inference for eigenvectors[J]. The Annals of Statistics, 1981, 9(4): 725-736.

[68] 茆诗松，王静龙，濮晓龙. 高等数理统计[M]. 北京：高等教育出版社, 1998.

[69] Lehmann E L. Testing Statistical Hypothesis[M]. New York: John Wiley, 1959.

[70] Cook R D, Wetzel N. Exploring Regression Structure With Graphics (with discussion)[J]. TEST, 1993, 2(1-2): 33-100.

[71] Cook R D, Critchley F. Identifying Regression Outliers and Mixtures Graphically[J]. Journal of the AmericanStatistical Association, 2000, 95(451): 781-794.

[72] Cook R D. Regression Graphics[M]. New York: John Wiley, 1998.

[73] Cook R D. On the interpretation of regression plots. Journal of the American Statistical Association[J]. 1994, 89(425): 177-189.

[74] Gervini D, Rousson V. Criteria for Evaluating Dimension-reducing Components for Multivariate Data[J]. The American Statistician, 2004, 58(1), 72-76.

[75] 陈希孺. 数理统计引论[M]. 北京: 科学出版社，1997.

[76] 黄廷祝, 钟守铭, 李正良. 矩阵理论[M]. 北京: 高等教育出版社,2003.

[77] 茆诗松, 王静龙, 史定华, 等. 统计手册[C]. 北京: 科学出版社, 2003.

[78] 严士健, 王隽骧, 刘秀芳. 概率论基础[M]. 北京: 科学出版社,1982.

[79] 张尧庭, 方开泰. 多元统计分析引论[M]. 北京: 科学出版社, 1982.

[80] 钟开来. 概率论教程[M]. 刘文, 吴让泉, 译, 李志阐, 校. 上海: 上海科学出版社, 1989.

[81] Lehmann E L. Theoery of point estimation[M]. NewYork : John Wiley, 1980.

[82] Kato T. Perturbation Theory for linear operation[M]. New York: Springer, 1983.

[83] Nolan D, Pollard D. U-process : rate of convergence[J]. The Annals of Statistics, 1991, 86(2), 316-324.

[84] Li K C. On principal hessian directions for data visualization and dimension reduc tion: another application of stein’s lemma[J]. Journal of the American Statistical Association, 1992, 87(420): 1025-1039.

[85] Cook R D. Principal Hessian direction revisited (with discussion)[J]. Journal of the American Statistical Association, 1998, 93(441): 84-100.

[86] Li K C, Duan N.Regression analysis under link violation[J]. The Annals of Statistics, 1989, 17(3): 1009-1052.

[87] Friedman J H, Stuetzle W. Projection pursuit regression[J]. Journal of the American Statistical Association, 1981, 30(376): 817-823.

[88] Huber P. Projection pursuit(with discussion)[J]. The Annals of Statistics, 1985, 13(2): 435-475.

[89] Cox D. Asymptotics for M-type smoothing spline[J]. The Annals of Statistics, 1983, 11(2): 530-551.

[90] Hall P, Patil P. Formulae for mean integrated squared error of nonliear wavelet-based density estimators[J]. The Annals of Statistics, 1995, 23(3): 905-928.

[91] Fan J, Gijbels I. Local Polynomial Modeling and Its Applications[M]. London, UK: Chapman and Hall, 1996.

[92] Xia Y, Tong H, Li W K, et al. An adaptive estimation of dimension reduction space (with Discussion)[J]. Journal of the Royal Statistical Society B, 2002, 64(2): 363-410.

[93] Hadle W, Sttoker T M. Investigating smooth multiple regression by the method of average derivatives[J]. Journal of the American Statistical Association, 1989, 84(408): 986-995.

[94] Bura E. Using linear smoothers to assess the structural dimension of regression[J]. Statistica Sinica, 2003, 13(1): 143-162.

[95] Powell J L, Stock G H, Stoker T M. Semiparametric estimation of index coefficients[J]. Econometica, 1989, 57(6): 1403-1430.

[96] Stone C J. An asymptotic optimal windows selection rule for kernel density estimates[J]. The Annals of Statistics, 1984, 12(4): 1285-1297.

[97] Cook R D, Li B. Dimension reduction for conditional mean in regression[J].The Annals of Statistics, 2002, 30(2): 455-474.

[98] Chiaromonte F, Cook R D, Li B. Sufficient dimension reduction in regression with categorical predictiors[J].The Annals of Statistics, 2002, 30(2): 475-497.

[99] Yin X, Cook R D. Dimension Reduction for the conditional kth moment regression[J]. Journal of the Royal Statistical Society B, 2002, 64(2), 159-175.

[100] Ma J J, Xu X Z, Zhao J L. Inverse regression in binary response LDV model[J]. Communications in Statistics-Theory and methods, 2008, 37(2): 233-246.

[101] Zhao J L, Xu X Z, Ma J J. Extending SAVE and PHD[J]. Communications in Statistics, Theory and methods, 2008, 36(8): 1591-1606.

[102] 田茂再. 高等分层分位回归建模理论[M]. 北京: 科学出版社, 2015.

[103] 马建军, 徐兴忠. 多项 Probit 模型中的回归系数的逆回归估计[J]. 应用概率统计, 2008 , 24 (5): 501-512.

[104] 马建军, 徐兴忠. 多项选择模型回归系数的估计[J]. 高校应用数学学报, 2008 , 23 (3): 000287-294

[105] 马建军, 徐兴忠. 多重二元响应 Probit 模型的渐近有效估计[J]. 生物数学学报, 2008, 23(4): 677-686.

反侵权盗版声明

举报电话：（010）88254396；（010）88258888
传　　真：（010）88254397
E-mail：　dbqq@phei.com.cn
通信地址：北京市万寿路 173 信箱
　　　　　电子工业出版社总编办公室
邮　　编：100036